

## Supplementary Tables and Figures

**Supp. Table 1: Summary of pairwise mean  $F_{ST}$  calculated within the global dataset**

Populations/Groups	Mean Pairwise $F_{ST}$
<i>Within AGVP overall</i>	0.019
<i>Regional</i>	
Within	
East <sup>*</sup>	0.021
West <sup>**</sup>	0.006
South <sup>§</sup>	0.001
Across	
West-East	0.027
West-South	0.014
South-East	0.027
<i>Linguistic</i>	
Within	
Niger-Congo <sup>†</sup>	0.009
Nilo-Saharan <sup>‡</sup>	NA
Afro-Asiatic <sup>¶</sup>	0.008
Khoe-San <sup>***</sup>	0.019
Across	
Niger-Congo – Nilo-Saharan	0.018
Nilo-Saharan – Afro-Asiatic	0.025
Afro-Asiatic – Niger-Congo	0.048
Niger-Congo – Khoe-San	0.053
Afro-Asiatic – Khoe-San	0.058
Nilo-Saharan – Khoe-San	0.058
<i>Global</i>	
Africa-Europe <sup>§</sup>	0.099
Africa-Asia <sup>§§</sup>	0.124
Africa-America	0.093

\*'East Africa' includes the populations Luhya, Kalenjin, Kikuyu, Baganda, Banyarwanda, Barundi, Oromo, Amhara and Somali

\*\* 'West Africa' includes the populations Fula, Jola, Wolof, Mandinka, Yoruba, Igbo and Ga-Adangbe

§ 'South Africa' includes the Zulu and Sotho ethno-linguistic groups

† Niger-Congo population groups include Jola, Fula, Wolof, Mandinka, Yoruba, Igbo, Ga-Adangbe, Luhya, Baganda, Banyarwanda, Barundi, Kikuyu, Zulu and Sotho

‡Nilo-Saharan populations include Kalenjin only

¶ Afro-Asiatic linguistic groups here include Oromo, Amhara and Somali from Ethiopia

\*\*\*Khoe-San populations include Ju/'hoansi, Khwe, !Xun, Nama, Karretjie, GuiGhanaKgal and Khomani

§ 'Europe' includes the populations CEU, GBR, FIN, IBS and TSI from the 1000 Genomes Project

§§ 'Asia' includes the populations CHB, CHS, CDX, KHV and JPT from the 1000 Genomes Project

**Supp. Table 2: *f*3 tests for CEU and Ju/'hoansi admixture in AGVP populations**

<b>pop A</b>	<b>pop B</b>	<b>pop C</b>	<b>f3,C;A,B</b>	<b>Z score</b>
CEU	YRI	Zulu	0.00543	33.5
CEU	YRI	Fula	-0.01115	-59.4*
CEU	YRI	Jola	0.005366	31.9
CEU	YRI	Mandinka	-4.5E-05	-0.3
CEU	YRI	Wolof	-0.00079	-5.4*
CEU	YRI	Baganda	-0.00316	-23.7*
CEU	YRI	Banyarwanda	-0.01206	-74.7*
CEU	YRI	Barundi	-0.00904	-64.1*
CEU	YRI	LWK	-0.00322	-21.5*
CEU	YRI	Kalenjin	-0.01042	-50.5*
CEU	YRI	Kikuyu	-0.01618	-89.3*
CEU	YRI	Somali	-0.01459	-47.4*
CEU	YRI	Amhara	-0.02385	-87.0*
CEU	YRI	Oromo	-0.02391	-85.2*
CEU	YRI	Igbo	0.000502	5.3
CEU	YRI	Ga-Adangbe	0.000008	0.1
CEU	YRI	Sotho	0.005022	31.3
YRI	Ju/'hoansi	Zulu	-0.00705	-44.6*
YRI	Ju/'hoansi	Sotho	-0.00984	-64.8*
YRI	Ju/'hoansi	Fula	0.005064	38.4
YRI	Ju/'hoansi	Jola	0.007087	51.0
YRI	Ju/'hoansi	Mandinka	0.003206	29.1
YRI	Ju/'hoansi	Wolof	0.005002	40.0
YRI	Ju/'hoansi	Baganda	-0.00067	-6.2*
YRI	Ju/'hoansi	Barundi	-0.00207	-19.4*
YRI	Ju/'hoansi	Banyarwanda	-0.00044	-3.8*
YRI	Ju/'hoansi	LWK	-0.00017	-1.4
YRI	Ju/'hoansi	Kikuyu	0.003081	20.3
YRI	Ju/'hoansi	Kalenjin	0.009667	49.4
YRI	Ju/'hoansi	Igbo	-0.0004	-5.4*
YRI	Ju/'hoansi	Ga-Adangbe	-5.9E-05	-0.8
YRI	Ju/'hoansi	Oromo	0.042493	92.4
YRI	Ju/'hoansi	Amhara	0.048456	99.4
YRI	Ju/'hoansi	Somali	0.043256	95.0

\* Statistically significant results for admixture

**Supp Table 3: *f*3 analysis for hunter-gatherer admixture using different hunter-gatherer reference populations**

Source1	Source2	Target	f3	SE	Z
MbutiPygmy	YRI	Igbo	-0.00055	0.00009	-6.072
MbutiPygmy	YRI	Ga-Adangbe	0.000237	0.000101	2.346
MbutiPygmy	YRI	Fula	0.006759	0.000172	39.366
MbutiPygmy	YRI	Jola	0.007632	0.00018	42.44
MbutiPygmy	YRI	Wolof	0.005733	0.000162	35.28
MbutiPygmy	YRI	Mandinka	0.003972	0.000147	27.023
MbutiPygmy	YRI	Baganda	-0.00117	0.000146	-8.025
MbutiPygmy	YRI	Banyarwanda	0.000615	0.00015	4.107
MbutiPygmy	YRI	Barundi	-0.0017	0.00014	-12.097
MbutiPygmy	YRI	LWK	-0.00051	0.000153	-3.326
MbutiPygmy	YRI	Kalenjin	0.011348	0.000242	46.882
MbutiPygmy	YRI	Kikuyu	0.005397	0.000189	28.549
MbutiPygmy	YRI	Oromo	0.04873	0.000583	83.645
MbutiPygmy	YRI	Amhara	0.054609	0.000611	89.361
MbutiPygmy	YRI	Somali	0.048789	0.000584	83.544
MbutiPygmy	YRI	Zulu	-0.00023	0.000166	-1.377
MbutiPygmy	YRI	Sotho	-0.0017	0.000166	-10.22
Juhoan_North	YRI	Igbo	-0.0004	0.000102	-3.882
Juhoan_North	YRI	Ga-Adangbe	0.000011	0.000113	0.102
Juhoan_North	YRI	Fula	0.005899	0.000189	31.181
Juhoan_North	YRI	Jola	0.007238	0.000186	38.933
Juhoan_North	YRI	Wolof	0.005411	0.000171	31.595
Juhoan_North	YRI	Mandinka	0.003445	0.000153	22.497
Juhoan_North	YRI	Baganda	-0.00052	0.000144	-3.619
Juhoan_North	YRI	Banyarwanda	0.000143	0.000158	0.909
Juhoan_North	YRI	Barundi	-0.00169	0.000143	-11.823
Juhoan_North	YRI	LWK	0.000006	0.000149	0.04
Juhoan_North	YRI	Kalenjin	0.01064	0.000245	43.412
Juhoan_North	YRI	Kikuyu	0.004031	0.000201	20.069
Juhoan_North	YRI	Oromo	0.045122	0.000593	76.101
Juhoan_North	YRI	Amhara	0.051333	0.000643	79.858
Juhoan_North	YRI	Somali	0.0461	0.000595	77.543
Juhoan_North	YRI	Zulu	-0.00811	0.000202	-40.23
Juhoan_North	YRI	Sotho	-0.01088	0.000189	-57.683
BiakaPygmy	YRI	Igbo	-0.00031	0.00007	-4.381
BiakaPygmy	YRI	Ga-Adangbe	0.000486	0.000077	6.304
BiakaPygmy	YRI	Fula	0.007349	0.000148	49.514
BiakaPygmy	YRI	Jola	0.00822	0.000137	59.787
BiakaPygmy	YRI	Wolof	0.006141	0.000123	49.871
BiakaPygmy	YRI	Mandinka	0.004464	0.00011	40.755
BiakaPygmy	YRI	Baganda	0.001228	0.000106	11.534
BiakaPygmy	YRI	Banyarwanda	0.003139	0.000125	25.082

BiakaPygmy	YRI	Barundi	0.000888	0.000105	8.425
BiakaPygmy	YRI	LWK	0.001951	0.000114	17.096
BiakaPygmy	YRI	Kalenjin	0.015217	0.000207	73.441
BiakaPygmy	YRI	Kikuyu	0.00811	0.000163	49.797
BiakaPygmy	YRI	Oromo	0.052322	0.000561	93.31
BiakaPygmy	YRI	Amhara	0.058146	0.000598	97.185
BiakaPygmy	YRI	Somali	0.052071	0.000567	91.914
BiakaPygmy	YRI	Zulu	0.002524	0.000126	20.08
BiakaPygmy	YRI	Sotho	0.001357	0.000128	10.601
Hadza	YRI	Igbo	0.000069	0.000085	0.806
Hadza	YRI	Ga-Adangbe	0.000396	0.000098	4.064
Hadza	YRI	Fula	0.002996	0.000149	20.123
Hadza	YRI	Jola	0.006887	0.000156	44.126
Hadza	YRI	Wolof	0.003776	0.000142	26.527
Hadza	YRI	Mandinka	0.002876	0.000132	21.765
Hadza	YRI	Baganda	-0.0008	0.00012	<b>-6.679</b>
Hadza	YRI	Banyarwanda	-0.00228	0.000132	<b>-17.281</b>
Hadza	YRI	Barundi	-0.00304	0.000118	<b>-25.841</b>
Hadza	YRI	LWK	-0.00134	0.000132	<b>-10.154</b>
Hadza	YRI	Kalenjin	0.003112	0.00019	16.382
Hadza	YRI	Kikuyu	-0.00107	0.00016	<b>-6.698</b>
Hadza	YRI	OROMO	0.032014	0.000503	63.685
Hadza	YRI	AMHARA	0.037464	0.00053	70.688
Hadza	YRI	SOMALI	0.033076	0.000478	69.211
Hadza	YRI	Zulu	0.003557	0.000155	22.887
Hadza	YRI	Sotho	0.002898	0.000152	19.066

Supp Table 3 represents the  $f_3$  test for AGVP populations using different hunter-gatherer reference populations. Tests with significantly negative  $f_3$  test Z scores are highlighted.

**Supp Table 4: Proportion of Eurasian ancestry in AGVP populations**

Population	Eurasian ancestry		
	F4 ratio test with masking		ADMIXTURE (k=18)
	F4 ratio test (Method 1)%	(Method 2) %	
ACB	11.19	12.10	10.73
Amhara	47.82	49.73	54.70
ASW	19.11	20.22	19.32
Baganda	2.86	3.50	1.91
BantuKenya	2.33	3.05	1.93
BantuSouthAfrica	0.00	0.00	0.45
Banyarwanda	8.86	9.72	9.07
Barundi	5.68	6.46	5.93
BiakaPygmy	0.00	0.00	0.00
Damara	1.16	1.74	0.83
Dinka_Hammer	0.00	0.00	0.00
Fula	11.46	12.39	10.79
Ga-Adangbe	0.15	0.72	0.44
Gana	1.39	1.95	0.41
Gui	0.62	1.22	1.20
Hadza	5.29	5.93	1.08
Haiom	4.17	4.87	4.37
Himba	0.00	0.41	0.28
Hoan	0.04	0.68	0.13
Igbo	0.00	0.48	0.02
Jola	0.58	1.18	0.08
Juhoan_North	0.00	0.00	0.03
Juhoan_South	0.00	0.00	0.23
Kalenjin	10.14	10.99	7.30
Kgalagadi	0.00	0.45	0.00
Khwe	2.41	3.04	3.56
Kikuyu	12.91	13.89	13.34
LWK	1.85	2.50	0.65
Mandinka	1.92	2.48	1.12
Mbukushu	0.00	0.00	0.14
MbutiPygmy	0.00	0.00	0.00
Nama	12.93	13.87	14.41
Naro	0.12	0.67	0.85
Oromo	43.62	45.39	50.82
Shua	3.95	4.60	4.46
Somali	37.05	38.72	45.70
Sotho	0.00	0.00	0.48
Taa_East	0.00	0.00	0.00
Taa_North	1.11	1.68	0.58
Taa_West	0.00	0.00	0.07
Tshwa	3.05	3.80	1.70
Tswana	0.00	0.00	0.00
Wambo	0.00	0.00	0.05
Wolof	3.44	4.04	2.26
Xuun	0.74	1.16	0.86
Yoruba	0.00	0.00	0.06
Zulu	0.00	0.00	0.21

Supp Table 4 represents the proportion of Eurasian ancestry estimated using unadmixed reference populations in the f4 test (Method 1), ancestry masked reference and test populations (Method 2) in f4 tests, and ADMIXTURE clustering (k=18)- see Supplementary Note 3 for further detail.

**Supp Table 5. Proportion of HG ancestry among SSA populations**

ADMIXTURE HG ancestry	F4 ratio with Juhoan North reference		F4 ratio with Mbuti Pygmy reference	
	Method 1	Method 2	Method 1	Method 2
ACB	0.02	0.02	0.02	0.03
Amhara	0.01	0.02	0.05	0.07
ASW	0.02	0.02	0.03	0.04
Baganda	0.03	0.06	0.07	0.11
BantuKenya	0.03	0.04	0.05	0.08
BantuSouthAfrica	0.08	0.13	0.16	0.20
Banyarwanda	0.03	0.06	0.07	0.08
Barundi	0.04	0.07	0.08	0.09
BiakaPygmy	0.70	0.50	0.51	0.74
Damara	0.08	0.11	0.12	0.16
Dinka_Hammer	0.03	0.00	0.00	0.00
Fula	0.01	0.00	0.00	0.00
Ga-Adangbe	0.02	0.02	0.02	0.03
Gana	0.54	0.59	0.62	0.90
Gui	0.92	0.88	0.89	1.00
Hadza	0.02	0.24	0.27	0.34
Haiom	0.60	0.57	0.62	0.88
Himba	0.08	0.10	0.11	0.15
Hoan	0.76	0.76	0.78	1.00
Igbo	0.02	0.02	0.02	0.03
Jola	0.02	0.00	0.00	0.00
Juhoan_North	0.99	0.96	1.00	1.00
Juhoan_South	0.98	0.95	0.96	1.00
Kalenjin	0.02	0.04	0.06	0.03
Kgalagadi	0.24	0.34	0.37	0.54
Khwe	0.40	0.38	0.42	0.60
Kikuyu	0.01	0.05	0.07	0.06
LWK	0.02	0.06	0.07	0.07
Mandenka	0.02	0.00	0.00	0.00
Mandinka	0.02	0.00	0.00	0.00
Mbukushu	0.07	0.11	0.12	0.16
Nama	0.63	0.62	0.66	0.94
Naro	0.96	0.93	0.94	1.00
Oromo	0.02	0.03	0.06	0.04
Shua	0.37	0.38	0.42	0.60
Somali	0.00	0.00	0.03	0.00
Sotho	0.08	0.20	0.23	0.32
Taa_East	0.82	0.82	0.84	1.00
Taa_North	0.95	0.92	0.93	1.00
Taa_West	0.90	0.90	0.91	1.00
Tshwa	0.50	0.51	0.56	0.78
Tswana	0.15	0.23	0.26	0.36
Wambo	0.06	0.08	0.08	0.11
Wolof	0.01	0.00	0.00	0.00
Xuun	0.79	0.76	0.79	1.00
Yoruba	0.01	0.00	0.00	0.00
Zulu	0.04	0.22	0.21	0.28
				0.30

Supp Table 5 represents the proportion of HG ancestry estimated using unadmixed reference populations in the f4 test (Method 1), ancestry masked reference and test populations (Method 2) in f4 tests, and ADMIXTURE clustering ( $k=18$ )- see Supplementary Note 4 for further detail. The estimates with Mbuti Pygmy reference individuals are greater than those with Ju/'hoan\_North as the reference, as discussed in Supp Note 4, capped at 1 for f4 ratios exceeding this with Mbuti Pygmy reference.

**Supp. Table 6 Top signals of differentiation among African populations that remained after masking of non-SSA ancestry**

Chr	Pos	AGVP $F_{ST}$ (unmasked)	AGVP $F_{ST}$ (masked)	Eur-Afr $F_{ST}$	Genes within 50kb	AF CEU	AF YRI	AF CHB	AF Ethiopia	AF Zulu	AF LWK	AF Kalenjin
12	16403381	0.17	0.12	0.13	<i>SLC15A5</i>	0.35	0.00	0.35	0.28	0.00	0.01	0.05
5	100807308	0.16	0.07	0.29	NA	0.54	0.01	0.14	0.40	0.01	0.03	0.14
15	75083494	0.15	0.08	0.36	<i>CSK,CPLX3,LMAN1L,CYP1A2,ULK3</i>	0.69	0.00	0.15	0.35	0.01	0.00	0.11
17	72508425	0.15	0.07	0.58	<i>CD300C,CD300LB,CD300A</i>	0.92	0.05	0.68	0.55	0.02	0.09	0.25
8	5973222	0.15	0.09	0.19	NA	0.39	0.00	0.37	0.29	0.00	0.01	0.09
2	151910669	0.15	0.08	0.06	NA	0.26	0.01	0.40	0.42	0.02	0.05	0.11
12	65398704	0.14	0.08	0.20	<i>FLJ41278,WIF1</i>	0.58	0.07	0.24	0.67	0.09	0.14	0.32
9	133823473	0.14	0.09	0.19	<i>FIBCD1</i>	0.50	0.02	0.40	0.41	0.09	0.07	0.22
5	24283625	0.14	0.08	0.31	NA	0.63	0.01	0.44	0.40	0.01	0.03	0.15
16	59005027	0.14	0.08	0.24	NA	0.47	0.00	0.19	0.27	0.00	0.01	0.05
6	108996748	0.13	0.08	0.27	<i>FOXO3</i>	0.72	0.05	0.82	0.50	0.13	0.17	0.33
6	108999287	0.13	0.08	0.27	<i>FOXO3</i>	0.72	0.05	0.82	0.50	0.13	0.17	0.33
6	108996963	0.13	0.08	0.27	<i>FOXO3</i>	0.71	0.05	0.82	0.50	0.13	0.17	0.33
12	19168301	0.13	0.08	0.36	NA	0.65	0.00	0.52	0.31	0.00	0.02	0.08
5	161762603	0.13	0.09	0.25	NA	0.48	0.00	0.02	0.32	0.01	0.01	0.08
4	81058566	0.13	0.08	0.53	<i>PRDM8</i>	0.80	0.01	0.39	0.45	0.03	0.02	0.13
4	23864809	0.13	0.10	0.05	<i>PPARGC1A</i>	0.22	0.02	0.20	0.34	0.02	0.04	0.23
4	23858476	0.13	0.09	0.05	<i>PPARGC1A</i>	0.22	0.02	0.30	0.34	0.02	0.04	0.23
2	158638267	0.13	0.08	0.42	<i>ACVR1</i>	0.79	0.05	0.92	0.57	0.10	0.13	0.30
2	106076261	0.13	0.08	0.14	<i>FHL2</i>	0.38	0.02	0.27	0.40	0.01	0.03	0.09
1	163345829	0.13	0.09	0.08	<i>NUF2</i>	0.28	0.04	0.61	0.40	0.04	0.11	0.43
2	173784658	0.13	0.08	0.29	<i>RAPGEF4</i>	0.69	0.04	0.82	0.49	0.04	0.18	0.33
8	116913074	0.12	0.08	0.24	NA	0.77	0.23	0.87	0.67	0.10	0.29	0.50
13	80975547	0.12	0.08	0.09	NA	0.26	0.00	0.20	0.27	0.00	0.02	0.04
19	42603776	0.12	0.08	0.40	<i>GRIK5,POU2F2,ZNF574,MIR4323</i>	0.82	0.12	0.89	0.61	0.09	0.14	0.35
3	27720363	0.12	0.09	0.32	<i>EOMES</i>	0.91	0.18	0.40	0.73	0.16	0.34	0.49
16	17127208	0.12	0.08	0.09	NA	0.25	0.00	0.64	0.32	0.03	0.03	0.11
15	75204026	0.12	0.07	0.27	<i>FAM219B,COX5A,RPP25,SCAMP2,MPI</i>	0.62	0.01	0.18	0.32	0.01	0.00	0.10

8	10852164	0.12	0.08	0.31	<i>MIR598,XKR6</i>	0.55	0.00	0.92	0.33	0.02	0.03	0.27
3	152604702	0.12	0.12	0.20	<i>P2RY1</i>	0.46	0.01	0.16	0.30	0.01	0.02	0.07
2	75453787	0.12	0.07	0.08	<i>TACR1</i>	0.31	0.00	0.25	0.35	0.03	0.03	0.11
16	5244275	0.12	0.09	0.34	NA	0.61	0.03	0.42	0.34	0.01	0.05	0.14
7	82451558	0.12	0.08	0.15	<i>PCLO</i>	0.42	0.03	0.77	0.33	0.02	0.09	0.31
7	82453708	0.12	0.08	0.15	<i>PCLO</i>	0.42	0.03	0.77	0.33	0.02	0.09	0.31
20	37447919	0.11	0.09	0.04	<i>PPP1R16B,ACTR5</i>	0.34	0.05	0.27	0.41	0.13	0.10	0.44
3	15333551	0.11	0.07	0.34	<i>SH3BP5,CAPN7</i>	0.09	0.76	0.27	0.26	0.83	0.67	0.38
9	138078679	0.11	0.07	0.15	NA	0.49	0.06	0.30	0.46	0.07	0.19	0.35
2	176047906	0.11	0.13	0.36	<i>MIR933,ATP5G3,ATF2</i>	0.68	0.01	0.49	0.36	0.09	0.05	0.19
17	4351345	0.11	0.12	0.00	<i>SPNS3</i>	0.14	0.25	0.78	0.00	0.11	0.17	0.24
4	85546826	0.11	0.12	0.01	<i>WDFY3,CDS1</i>	0.04	0.04	0.01	0.02	0.35	0.07	0.10
7	82422937	0.11	0.09	0.10	<i>PCLO</i>	0.33	0.03	0.60	0.31	0.01	0.08	0.30
5	131421694	0.11	0.09	0.05	<i>IL3,CSF2</i>	0.37	0.04	0.33	0.45	0.11	0.16	0.42
10	21506018	0.11	0.09	0.36	<i>NEBL</i>	0.13	0.87	0.02	0.37	0.82	0.58	0.41
3	118481063	0.11	0.08	0.12	NA	0.45	0.04	0.30	0.38	0.05	0.15	0.33
5	128743299	0.11	0.07	0.05	NA	0.17	0.00	0.15	0.25	0.02	0.02	0.05
11	107756315	0.11	0.14	0.01	<i>SLC35F2,RAB34</i>	0.05	0.00	0.09	0.21	0.00	0.03	0.05
5	131508560	0.11	0.08	0.04	<i>P4HA2</i>	0.35	0.04	0.34	0.45	0.11	0.16	0.42
16	86114253	0.11	0.07	0.21	NA	0.45	0.02	0.29	0.25	0.00	0.02	0.11
1	201913763	0.11	0.11	0.21	<i>LMOD1,RNPEP,TIMM17A</i>	0.71	0.14	0.29	0.62	0.13	0.22	0.46
9	8242654	0.11	0.09	0.16	NA	0.25	0.00	0.30	0.20	0.00	0.00	0.01
4	25589755	0.11	0.10	0.12	NA	0.36	0.01	0.09	0.38	0.05	0.05	0.15
11	63357329	0.11	0.10	0.27	<i>HRASLS2,PLA2G16,ATL3,RARRES3</i>	0.75	0.07	0.61	0.58	0.21	0.12	0.25
17	237667	0.11	0.08	0.25	<i>RPH3AL,C17orf97</i>	0.62	0.05	0.40	0.41	0.02	0.03	0.14
5	131438682	0.11	0.08	0.05	<i>IL3,CSF2</i>	0.36	0.04	0.33	0.45	0.11	0.17	0.42

AGVP  $F_{ST}$  masked and unmasked represent the  $F_{ST}$  calculated across all 16 African populations when non-SSA ancestry was masked, and when this was left unmasked, respectively. Eur-Afr  $F_{ST}$  represents differentiation between Europe and Africa at these loci.

**Supp. Table 7. Differentiated loci with patterns of differentiation related to exposure to disease**

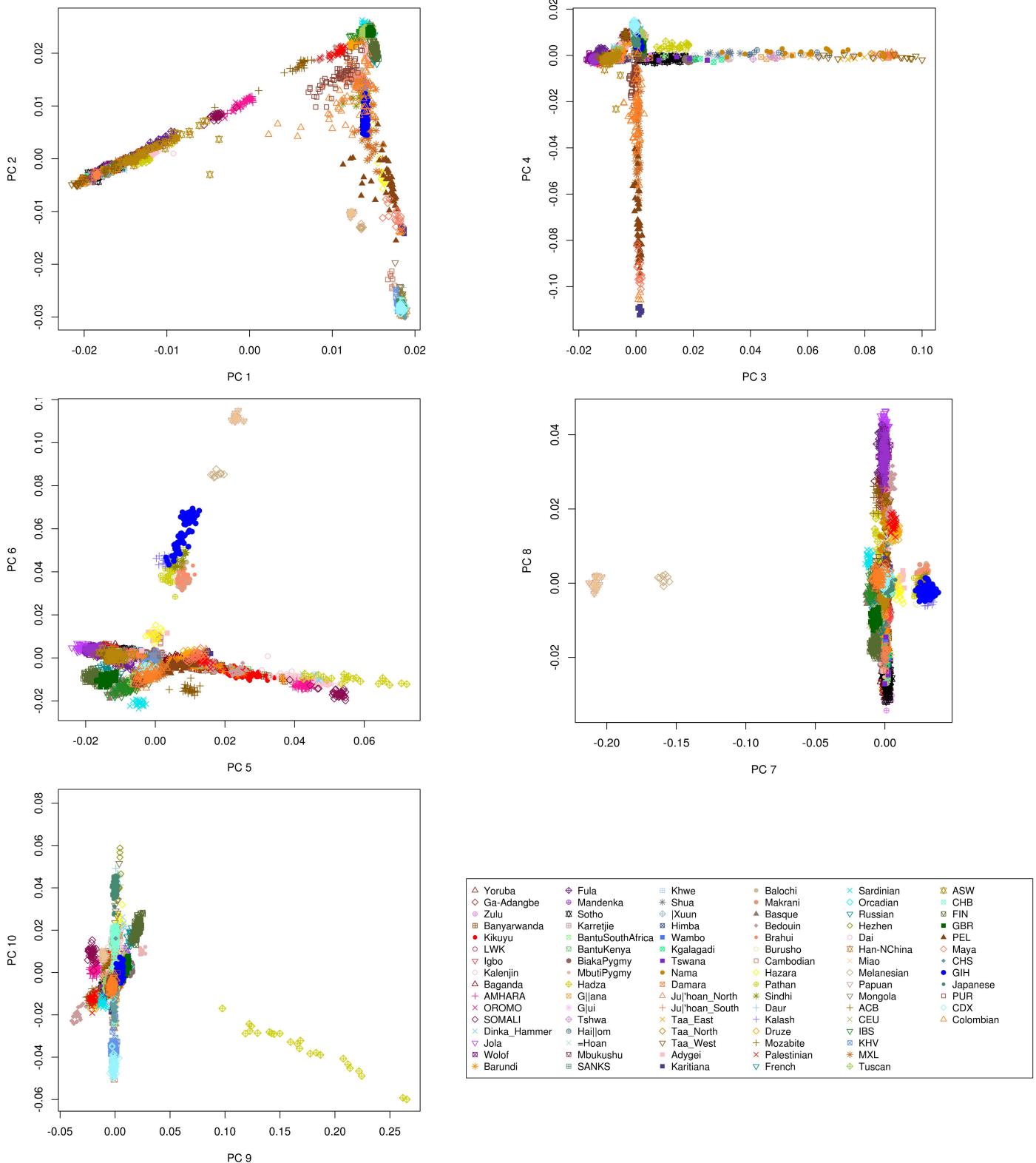
Exposure	ID	Anc	Der	GN	Jola	Ga-													
						Adangbe	YRI	Igbo	Baganda	Oromo	Amhara	Somali	Kikuyu	Kalenjin	LWK	Zulu	CEU	CHB	
<b>Malaria</b>	1:155258752	A	G	PKLR	0.33	0.35	0.36	0.31	0.18	0.48	0.49	0.47	0.29	0.22	0.27	0.10	0.64	0.24	
	11:5236954	C	T	HBB	0.33	0.28	0.30	0.29	0.29	0.31	0.33	0.26	0.20	0.19	0.32	0.20	0.36	0.03	
	1:25294878	C	T	RUNX3	0.89	0.79	0.76	0.73	0.74	0.50	0.56	0.47	0.55	0.45	0.72	0.68	0.50	0.29	
	12:7694017	G	A	CD163	0.30	0.34	0.40	0.38	0.26	0.06	0.04	0.12	0.18	0.06	0.26	0.21	0.06	0.38	
	1:206995478	G	A	IL10	0.49	0.50	0.49	0.54	0.72	0.85	0.77	0.77	0.79	0.76	0.70	0.77	0.78	0.36	
	1:196651832	C	T	CFH CD28-ICOS- CLTA4	0.91	0.96	0.92	0.91	0.95	1.00	1.00	1.00	0.98	0.99	0.95	0.99	1.00	1.00	
<b>Lassa fever</b>	2:204775855	G	T	LARGE	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.03	0.06	0.01	0.09	0.00	0.00	
	22:34094750	G	T	HLA-DRB1	0.14	0.19	0.31	0.22	0.09	0.00	0.00	0.00	0.03	0.07	0.11	0.07	0.00	0.00	
	6:32595194	C	T	DC-SIGNR	0.77	0.43	0.47	0.32	0.58	0.58	0.59	0.68	0.73	0.64	0.51	0.49	0.38	0.43	
	19:7836250	C	A	CXCR6	0.32	0.17	0.09	0.10	0.22	0.35	0.33	0.22	0.30	0.27	0.18	0.24	0.42	0.66	
	3:45886322	C	A	RNASEL	0.46	0.52	0.55	0.51	0.38	0.00	0.01	0.01	0.25	0.19	0.29	0.27	0.02	0.00	
	1:182549019	G	A	IFIH1	0.29	0.31	0.42	0.41	0.28	0.08	0.14	0.14	0.18	0.14	0.24	0.19	0.06	0.05	
<b>Trypanosomiasis</b>	2:163132687	A	C	LSECtin	0.06	0.06	0.11	0.19	0.03	0.02	0.00	0.03	0.04	0.02	0.03	0.05	0.00	0.00	
	19:7786250	C	A	APOL1	0.32	0.17	0.09	0.10	0.22	0.35	0.38	0.33	0.30	0.27	0.18	0.24	0.42	0.67	
	22:36601199	C	T	FASL	0.42	0.48	0.41	0.35	0.62	0.58	0.48	0.60	0.65	0.68	0.64	0.54	0.35	0.03	
	1:172617369	C	T	FAS	0.12	0.06	0.06	0.07	0.31	0.04	0.12	0.18	0.22	0.19	0.28	0.22	0.00	0.00	
	10:90778967	C	T	IL23R	0.42	0.44	0.42	0.37	0.28	0.65	0.37	0.35	0.42	0.48	0.36	0.36	0.38	0.55	
	1:67604219	T	C	SIGLEC6	0.16	0.12	0.14	0.12	0.36	0.38	0.39	0.46	0.40	0.43	0.30	0.17	0.27	0.32	
<b>Trachoma</b>	19:52031316	T	G	ABCA1	0.07	0.06	0.06	0.08	0.03	0.13	0.17	0.17	0.09	0.08	0.11	0.15	0.05	0.10	
	3:45936322	C	A	CXCR6	0.46	0.52	0.55	0.51	0.38	0.00	0.01	0.01	0.25	0.19	0.29	0.27	0.02	0.00	

Supp Table 7 represents the allele frequency distribution of derived alleles at differentiated loci among different populations. These loci show differentiation between populations thought to be in endemic vs non-endemic regions of exposure to specific infectious disease.

**Supp. Table 8: Number of variants with LD  $r^2 > 0.8$  with causal variant at specific loci in different populations**

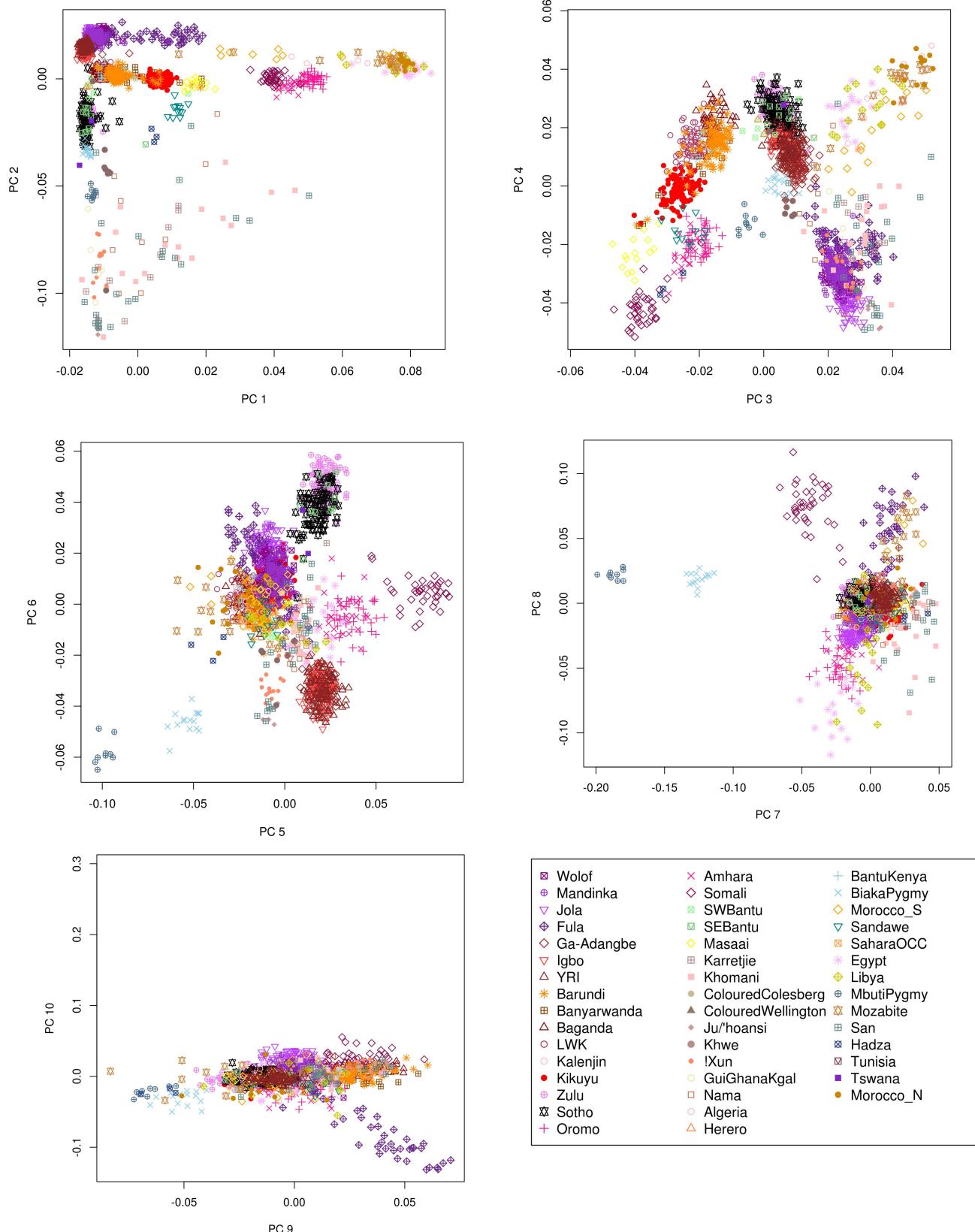
Locus	SNP	Europeans	Asians	Africans
<i>SORT1</i>	rs12740374	11	14	1
<i>TCF7L2</i>	rs7903146	10	4	0
<i>APOE</i>	rs429358	0	9	0
<i>HBB</i>	rs334	NA	NA	0
<i>APOL1</i>	rs73885319	NA	NA	4
<i>PRDM9</i>	rs6889665	28	28	10

**Supp Figure 1: Principal component analysis of global populations including the 1000 Genomes project, AGVP populations, and Human origins array.**



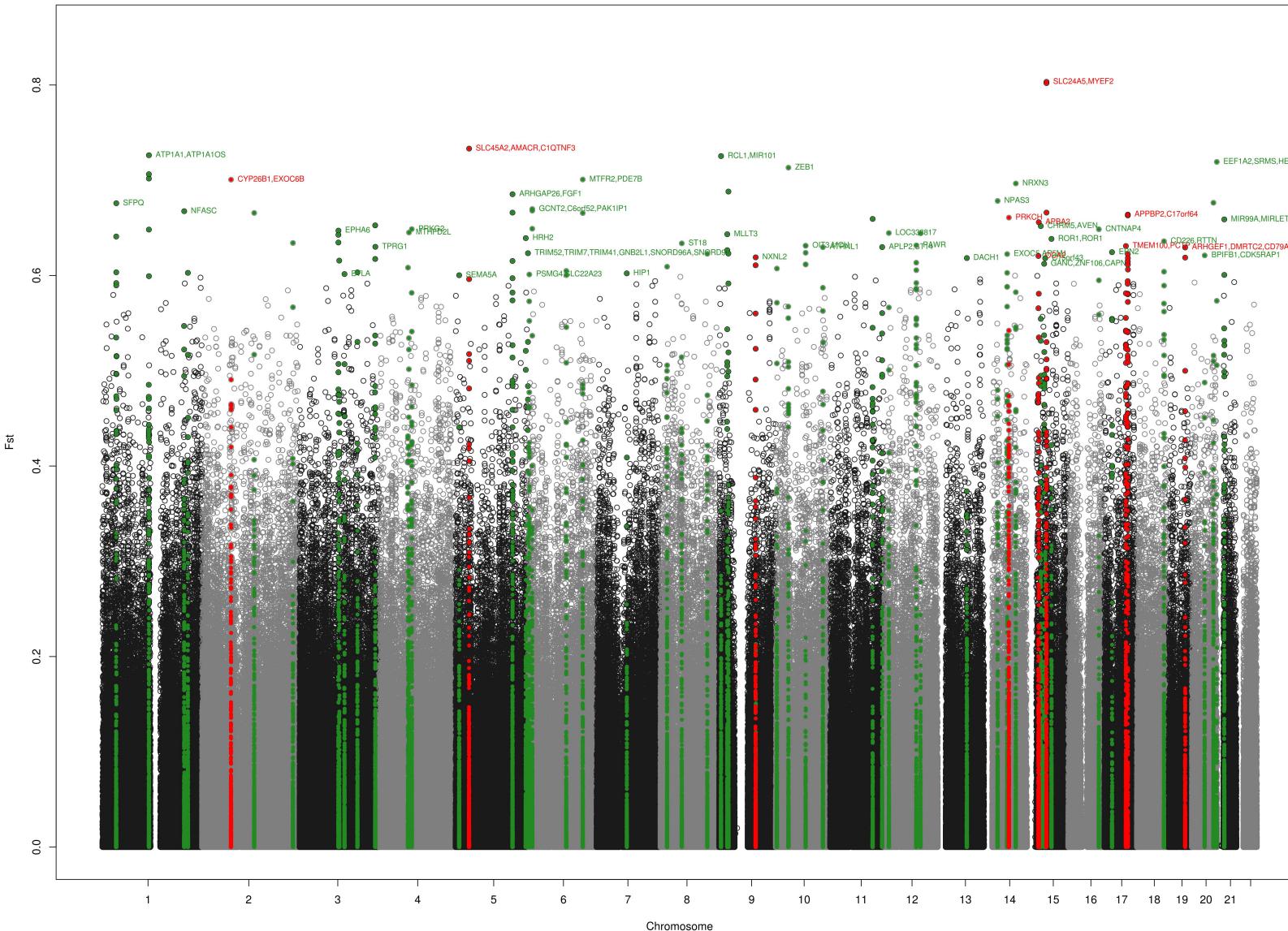
Supp. Figure 1 represents the first ten PCs for populations from the 1000 Genomes Project, AGVP and Human Origins data. PC 1 shows a cline among several African populations extending towards European populations, most likely to represent non-SSA gene flow. PC2 shows a clear separation between European and Asian populations. Subsequent PCs show more detailed structure between populations globally, and within African populations.

**Supp. Figure 2: PCA for AGV populations in the context of HGDP, 1000G, North African and Khoe-San populations**



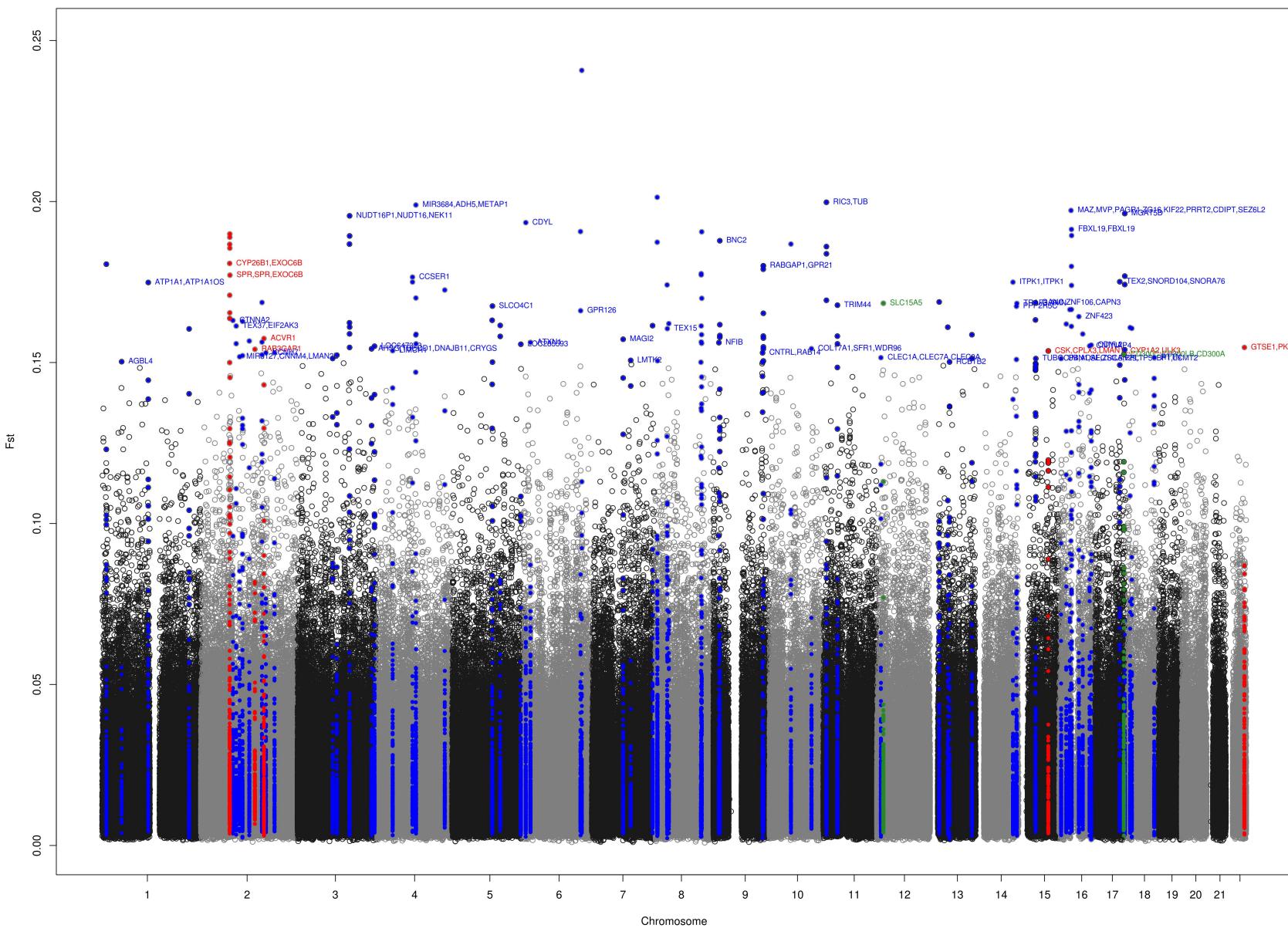
Supp. Figure 2 represents the first ten PCs for the African extended dataset, including AGVP populations, African populations from the 1000 Genomes Project, North African and Khoe-San population groups. PC 1 shows a cline among several African, most likely to represent non-SSA gene flow. PC2 shows a clear separation between Kho-San and other African populations, with South African populations showing a cline extending towards Kho-San groups. Subsequent PCs show more detailed structure between, and within African populations.

**Supp. Figure 3: Single SNP Wright's  $F_{ST}$  for Africa-Europe differentiation**



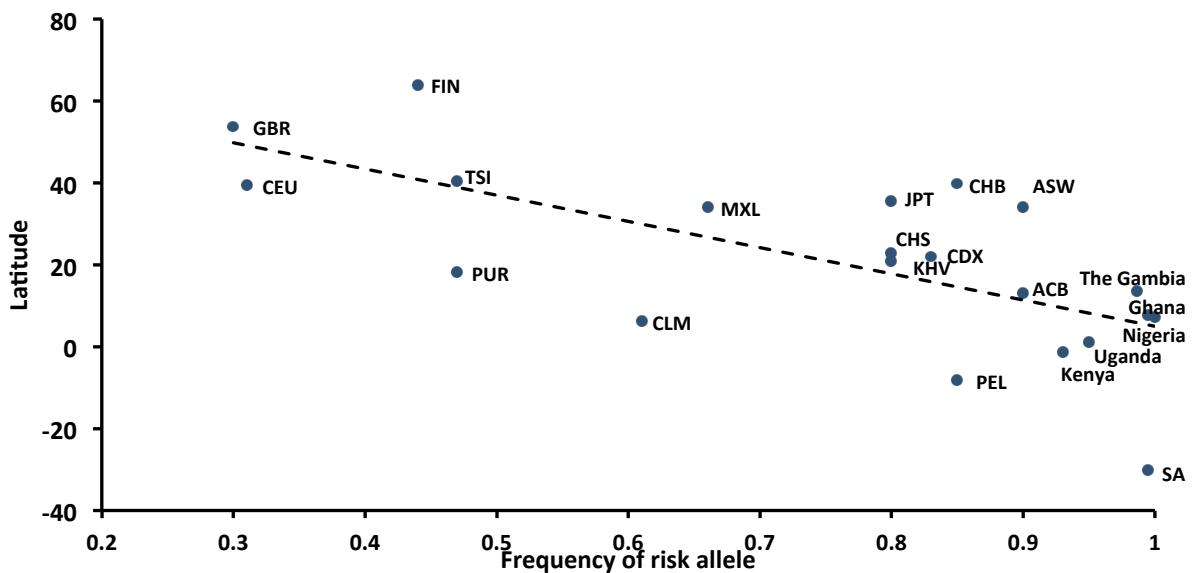
**Supp Figure 3.**  
 Signals with  $F_{ST}>0.60$  ( $n=99$ ) are annotated on the plot. Red signals represent those within genes previously noted to be under selection. Other signals are annotated in green. All signals shown here remained within the top 0.1% even after masking of non-SSA ancestry among African populations.

**Supp. Figure 4:**  $F_{ST}$  calculated across African populations



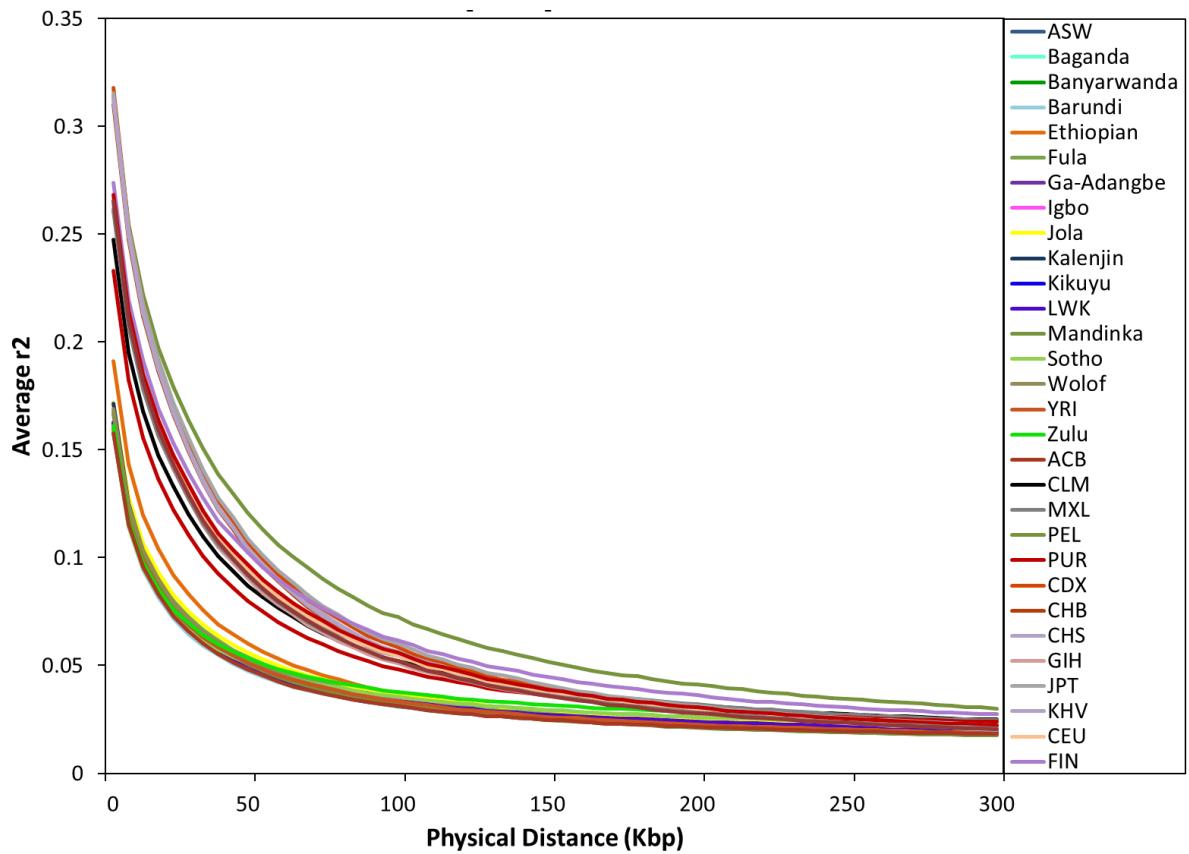
**Supp Figure 4.** Signals with  $F_{ST}>0.15$  ( $n=141$ ) are annotated on the plot. Red signals represent those within genes previously noted to be under selection. Signals that remained after masking of non-SSA ancestry are annotated in green.

**Supp. Figure 5: Correlation of risk allele frequency of the rs1378940 variant in the *CSK* region with latitude**



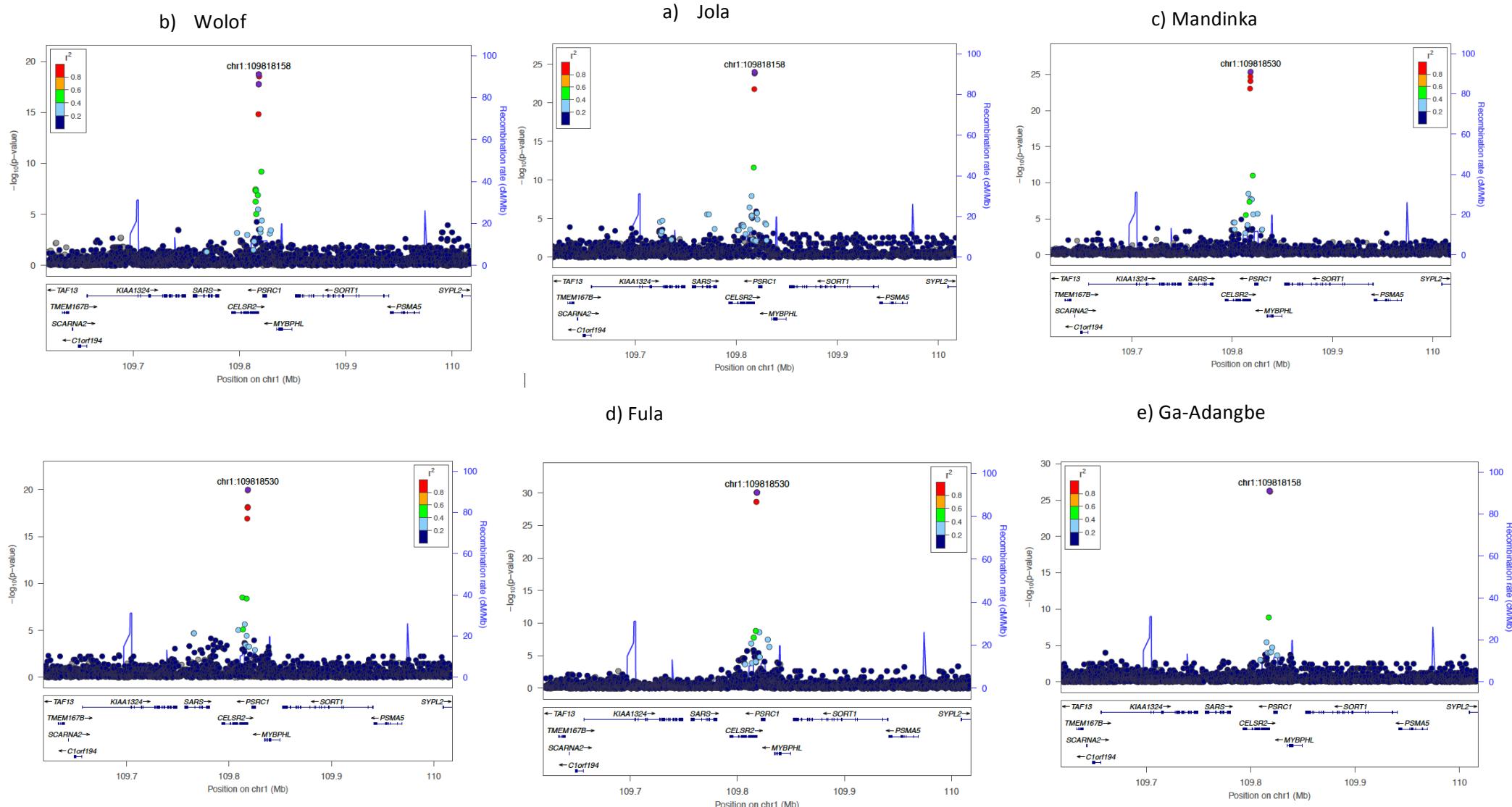
Supp Figure 5 shows the correlation between the frequency of the risk allele for the SNP rs1378940 associated with hypertension and latitude. This locus was found to be within the top 0.1% of differentiated loci within Africa, and within the top 1% of differentiated loci globally when examined across 33 global populations. There was a strong correlation between the frequency of the risk allele for hypertension and latitude across populations (Pearson's correlation=-0.67, p<0.0008). Differentiation at such loci associated with hypertension may reflect local evolutionary adaptation to temperature and latitude among different populations, which may have led to differential susceptibility to hypertension in these populations.

**Supp. Figure 6: LD decay among AGVP populations in a global context**

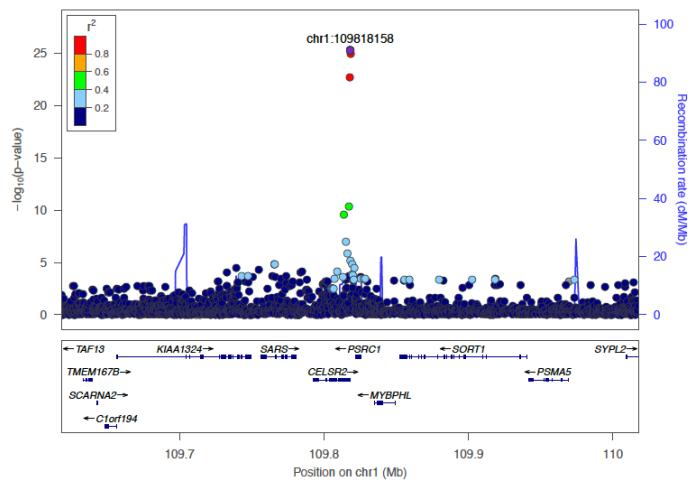


Supp. Figure 6 shows LD decay among global populations. Overall, African populations have a more rapid LD decay by physical distance as compared to other population groups. Among African populations, most populations have similar LD decay patterns, except Ethiopian populations, which seem to have a slower LD decay. This pattern is consistent with substantial non-SSA ancestry among Ethiopian populations.

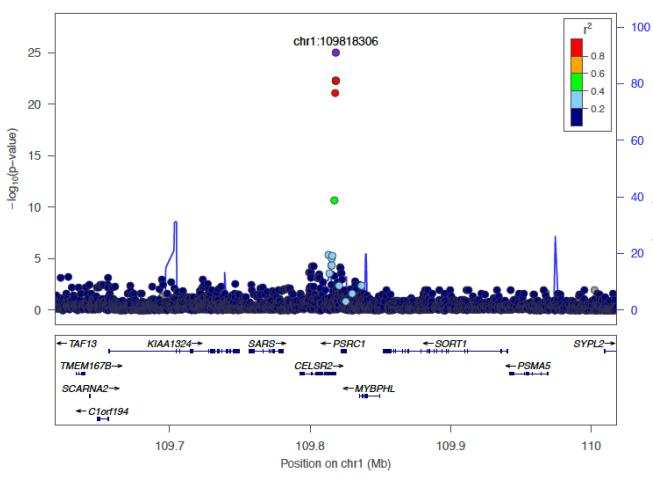
**Supp. Figure 7 Regional association plots and LD structure at the *SORT1* locus**



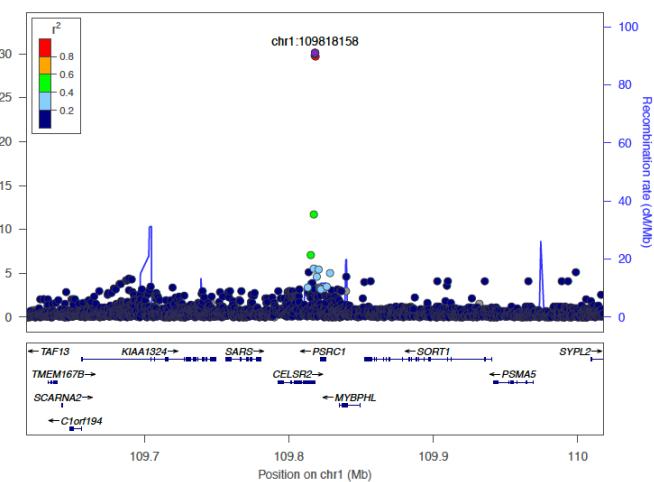
g) Igbo



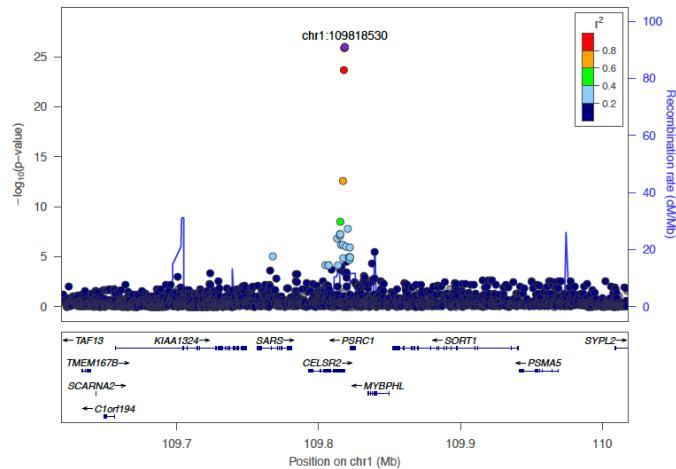
h) LWK



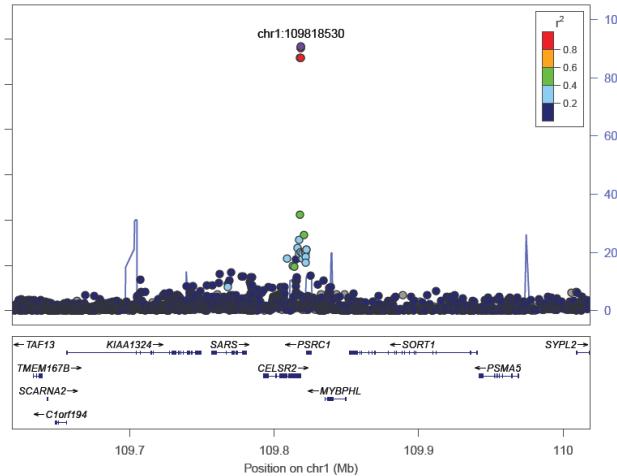
i) Kalenjin



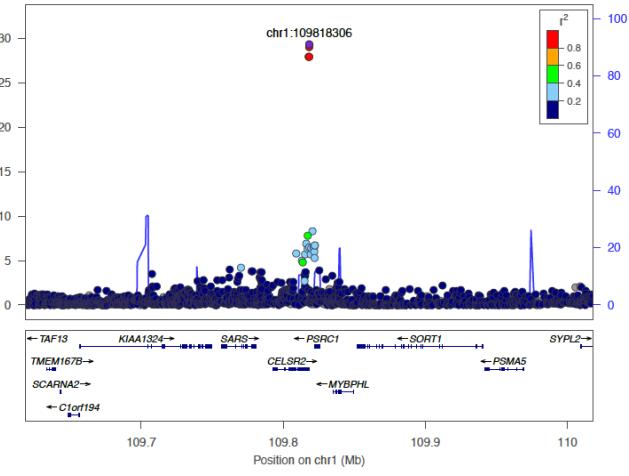
j) Kikuyu

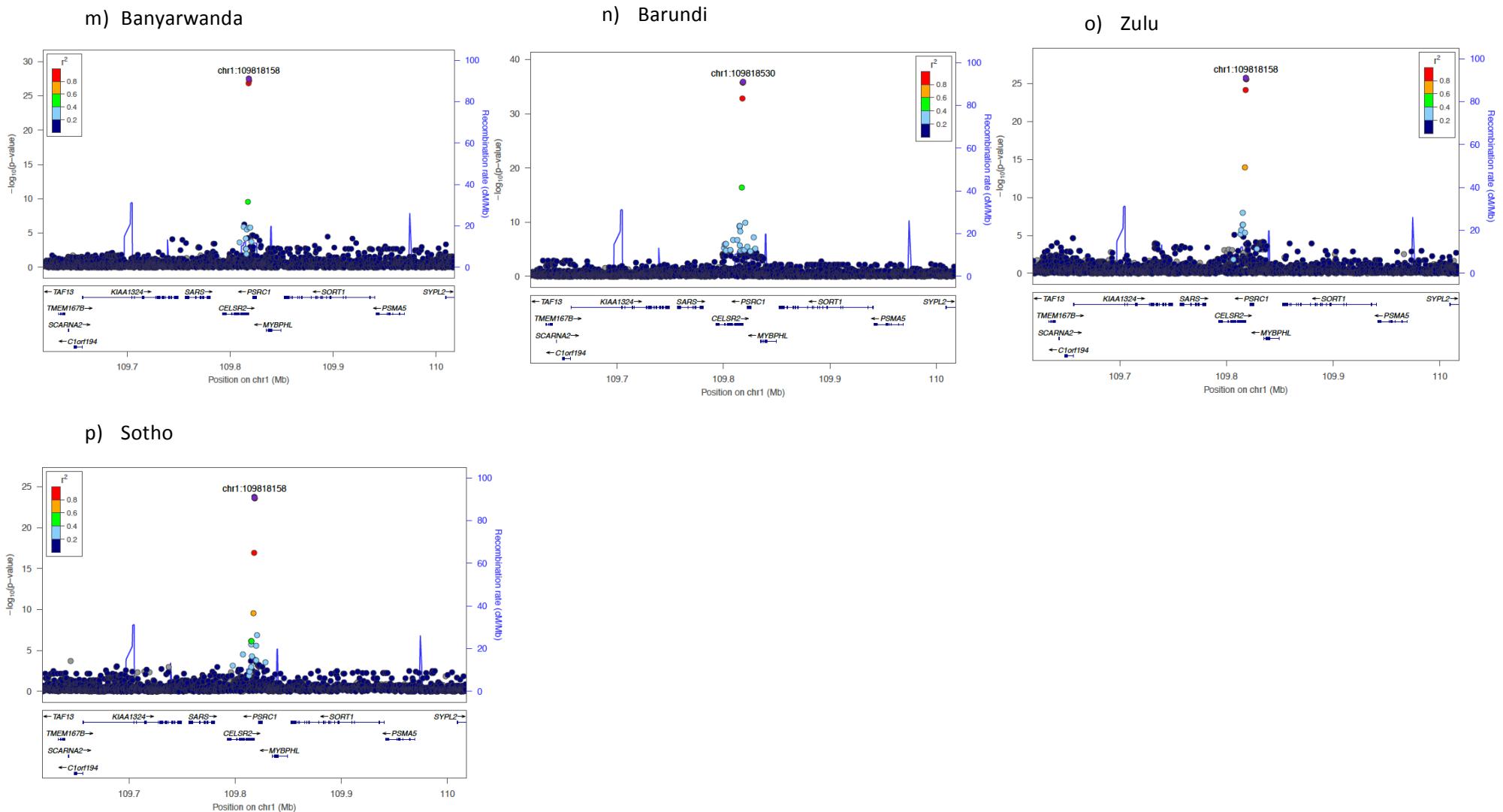


k) Ethiopia



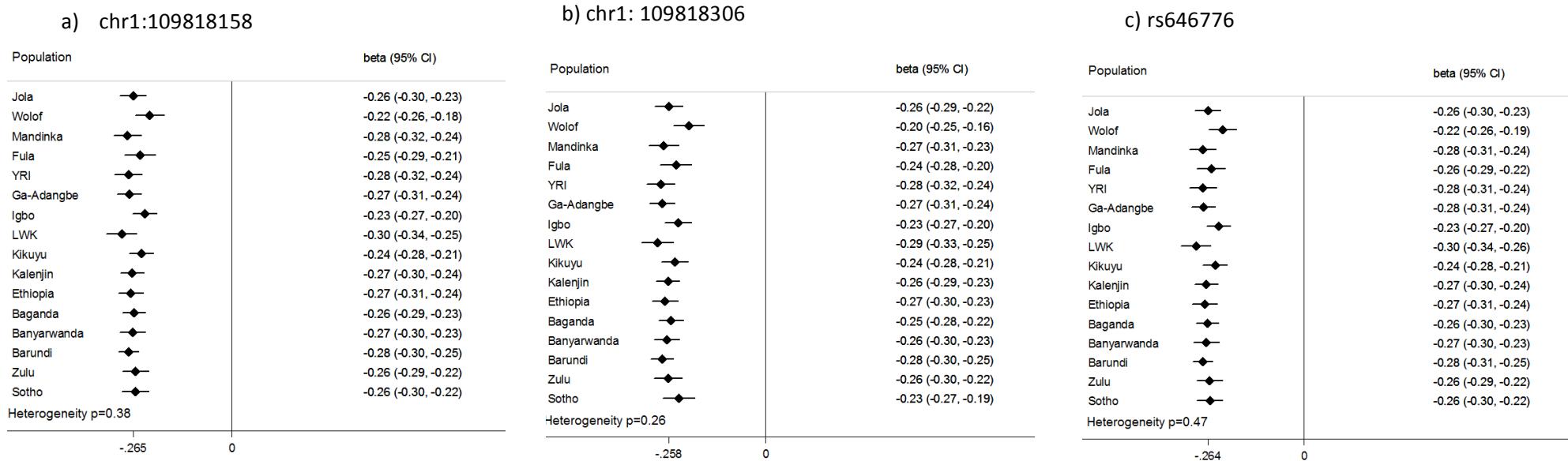
l) Baganda





Supp Figure 7 shows the pattern of association for a simulated trait around the *SORT1* locus in different African populations. Regional association plots show a similar pattern of association, suggesting homogenous LD structure across African populations in this region.

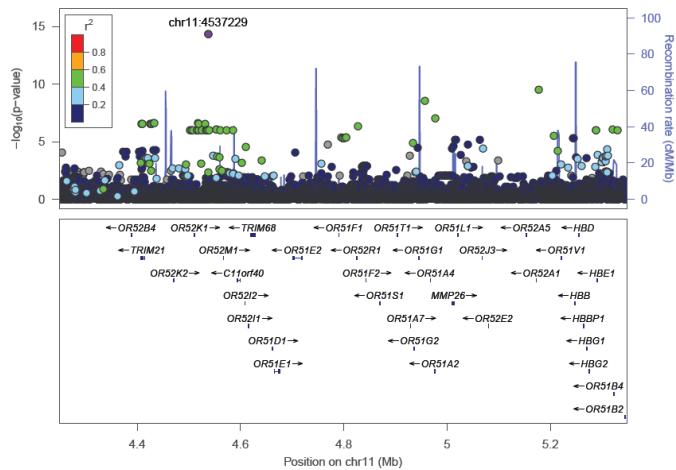
**Supp. Figure 8: Homogeneity in effect sizes in different African populations at the *SORT1* locus**



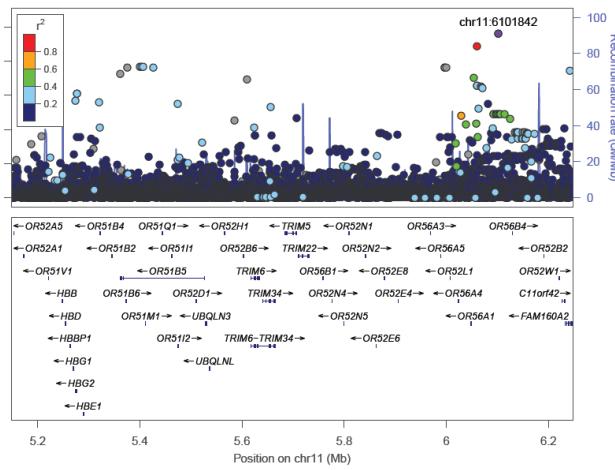
Supp. Figure 8 represents association estimates for the most significantly associated SNPs with a trait simulated with the causal *SORT1* variant. Although different top candidates seem to be observed in some populations, the association estimates are homogeneous across all populations, suggesting similar LD patterns across populations, and reproducibility of signals at this locus in association analyses.

**Supp. Figure 9: Regional association plots and LD structure at the sickle cell locus for a simulated trait**

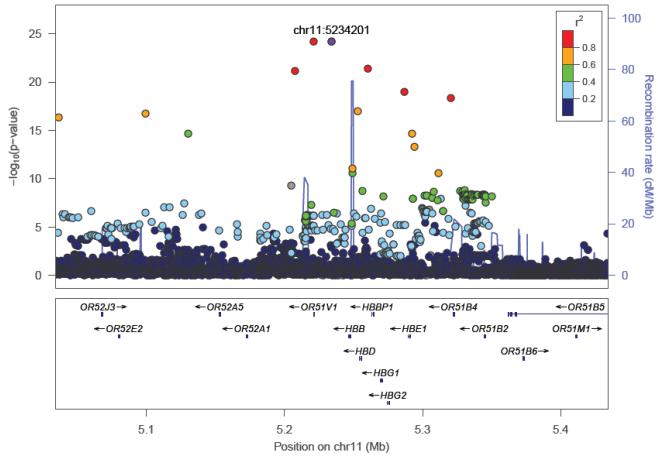
a) Mandinka



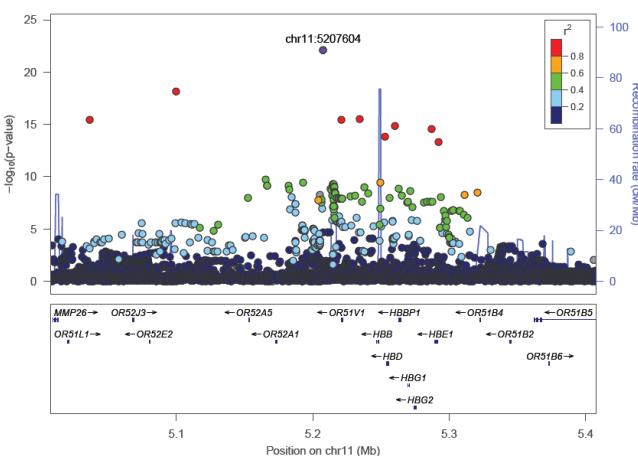
a) Wolof



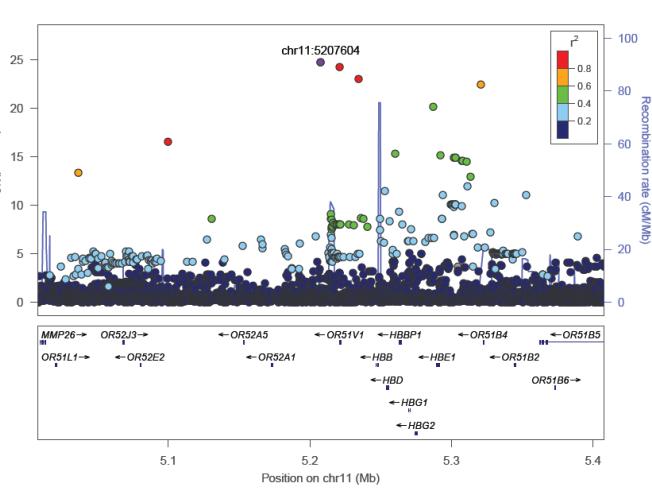
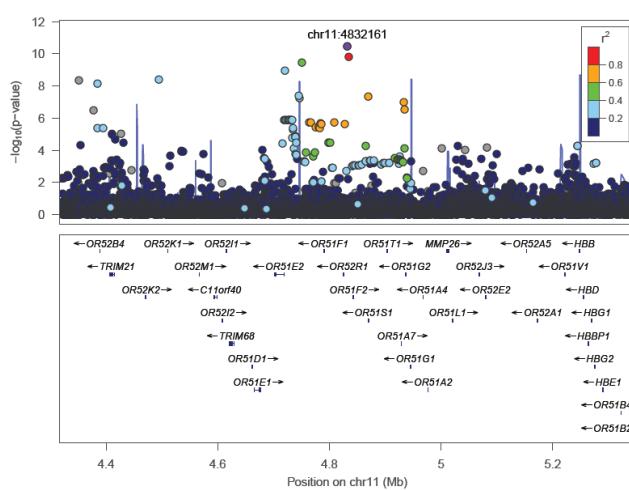
c) Ga-Adangbe

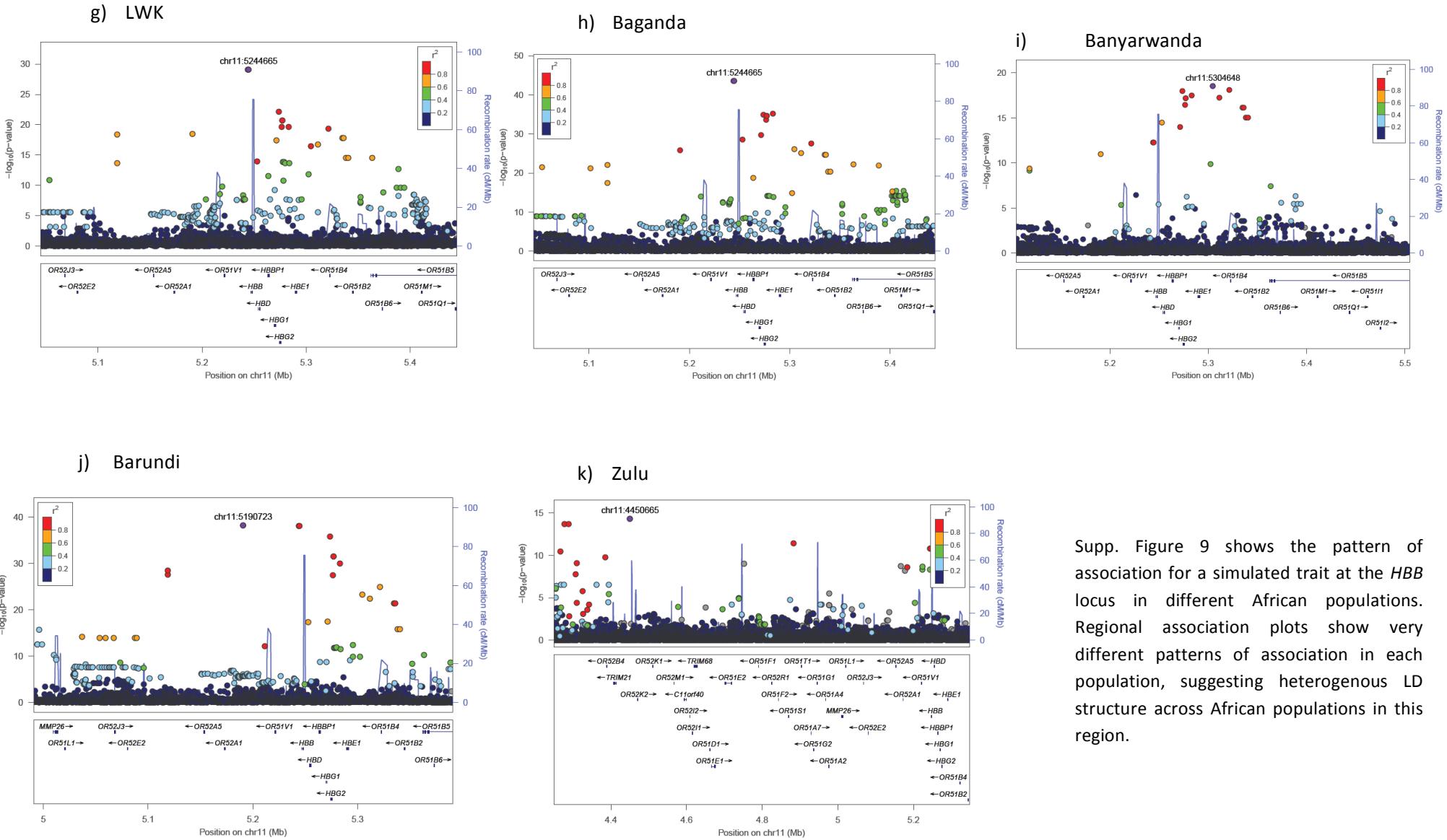


e) YRI



f) Kalenjin

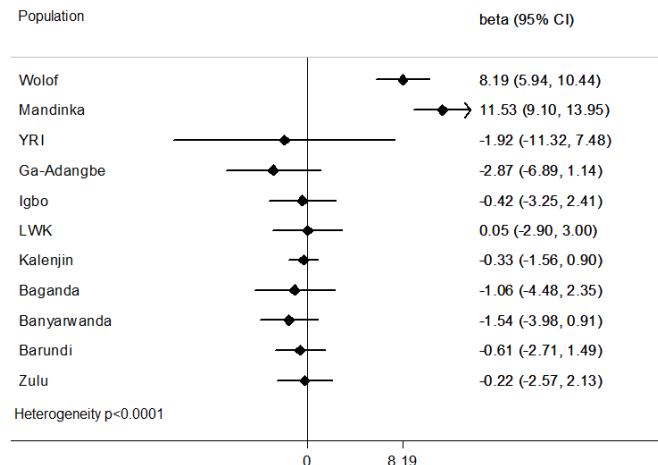




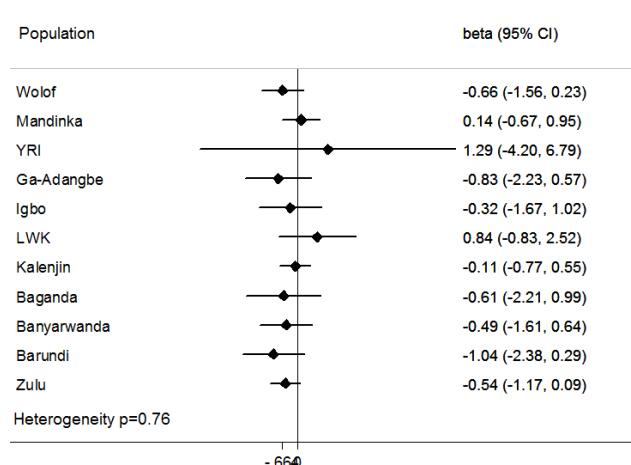
Supp. Figure 9 shows the pattern of association for a simulated trait at the *HBB* locus in different African populations. Regional association plots show very different patterns of association in each population, suggesting heterogenous LD structure across African populations in this region.

**Supp. Figure 10: Heterogeneity in effect size for sickle cell locus among different populations**

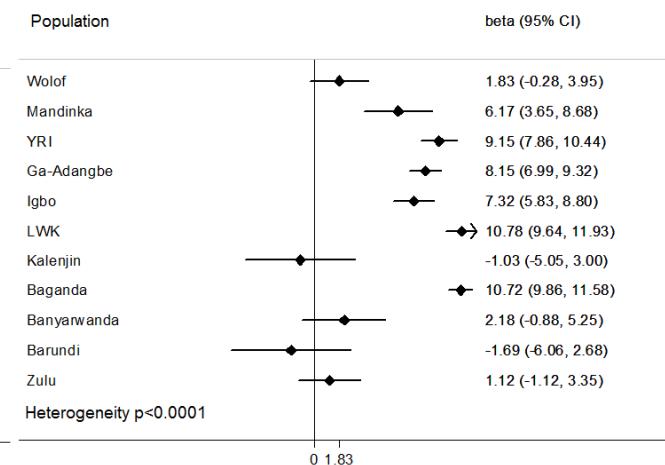
a) rs186588870



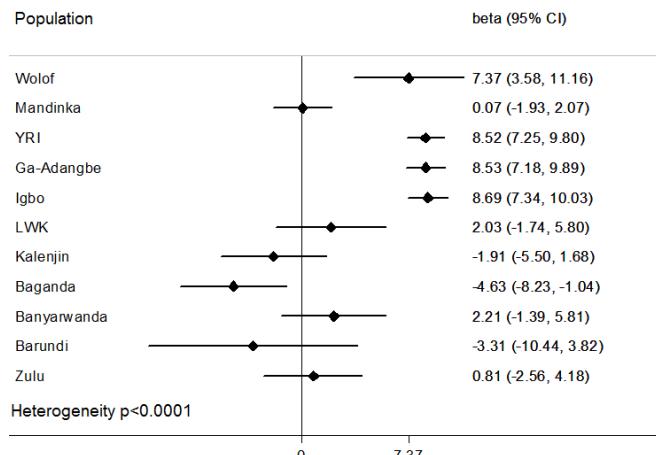
b) rs150301283



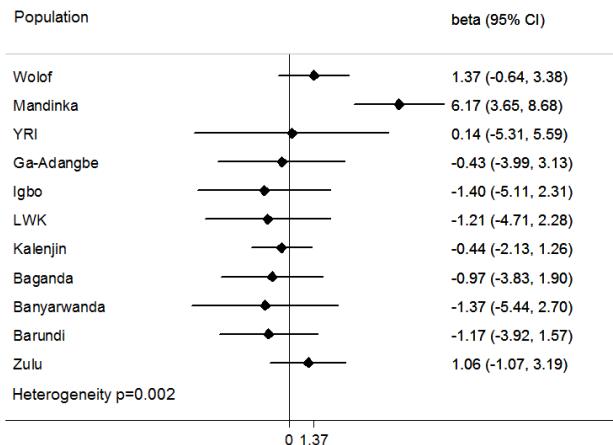
c) chr11:5207604



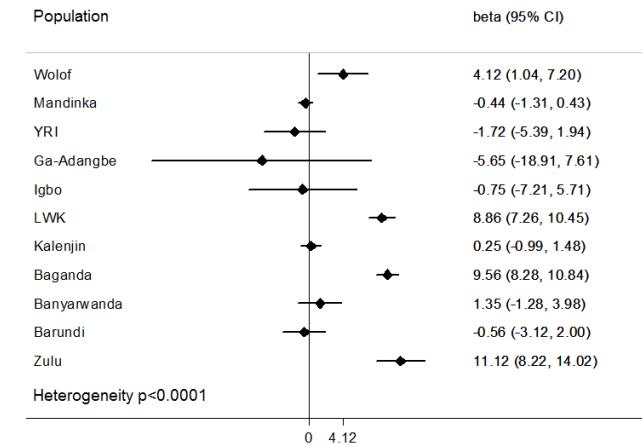
d) rs73400658



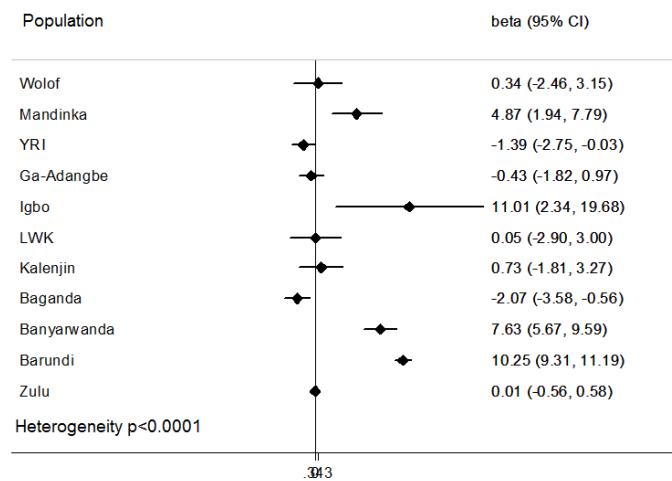
e) rs141845179



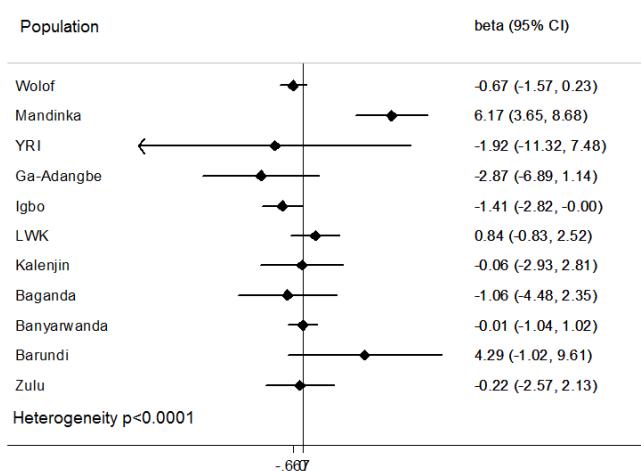
f) rs150545194



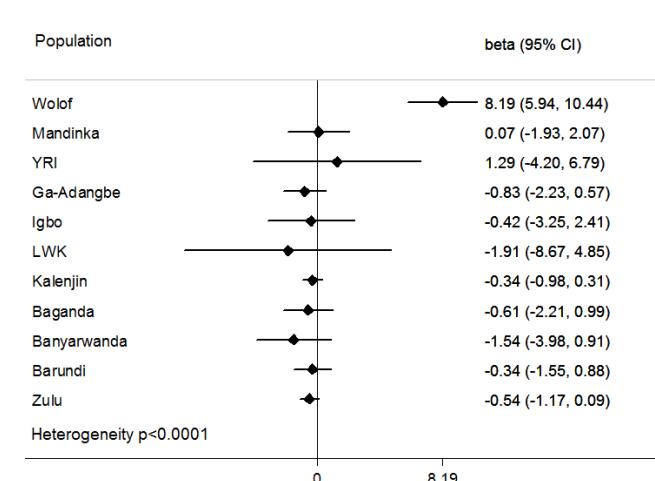
g) kgp12988299



h) rs142667079

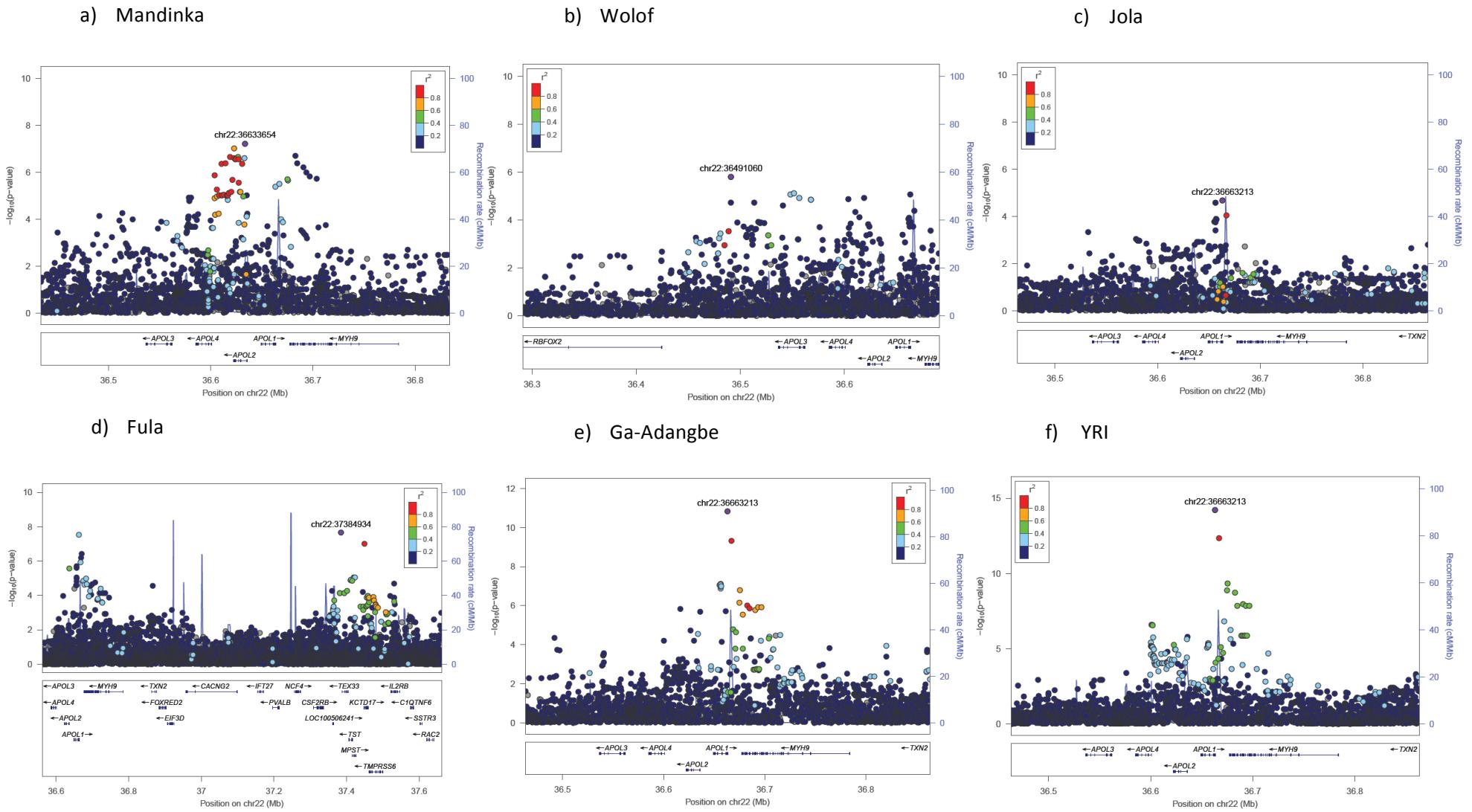


i) rs186786216

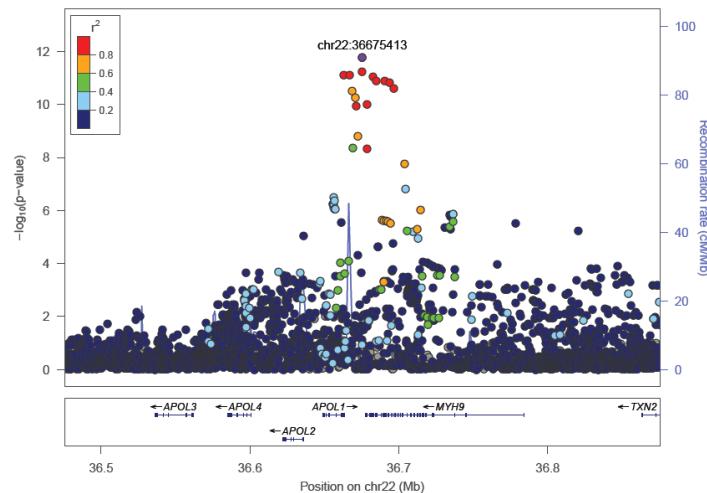


Supp. Figure 10 represents association estimates for the most significantly associated SNPs with a trait simulated with the rs334 SNP at the sickle cell locus. The most significant associated candidate after masking the causal variant seem to be very different across all populations. Moreover, the association estimates for many of these candidates are markedly heterogeneous across all populations, suggesting different LD patterns across populations.

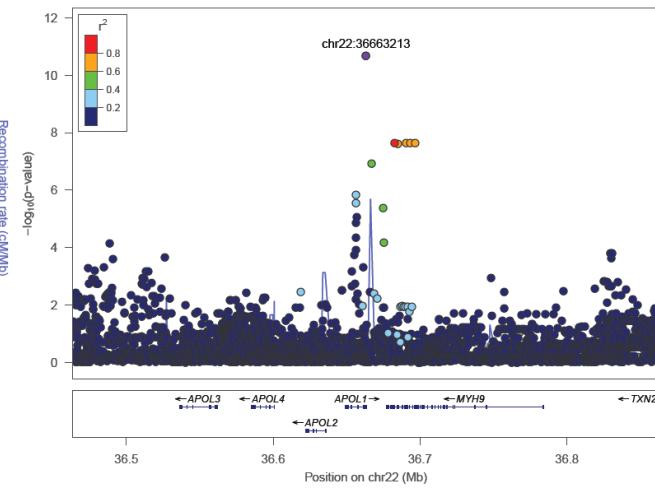
**Supp. Figure 11: Regional association plots and LD structure at the *APOL1* locus**



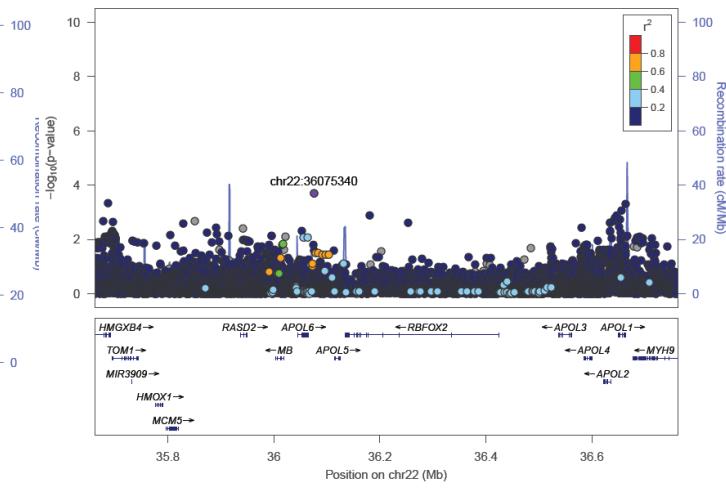
g) Igbo



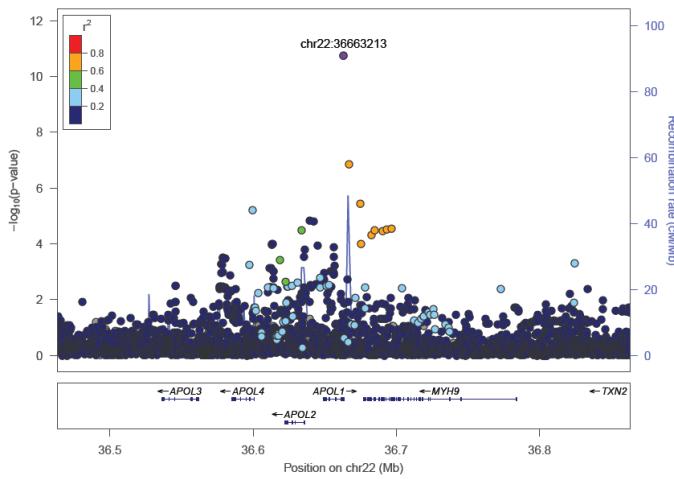
h) LWK



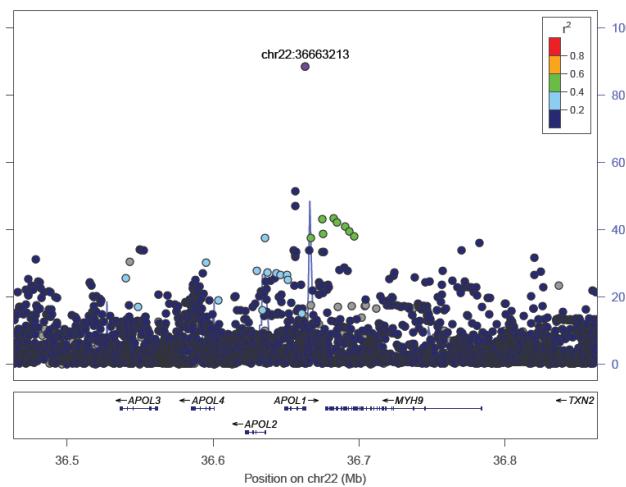
i) Kalenjin



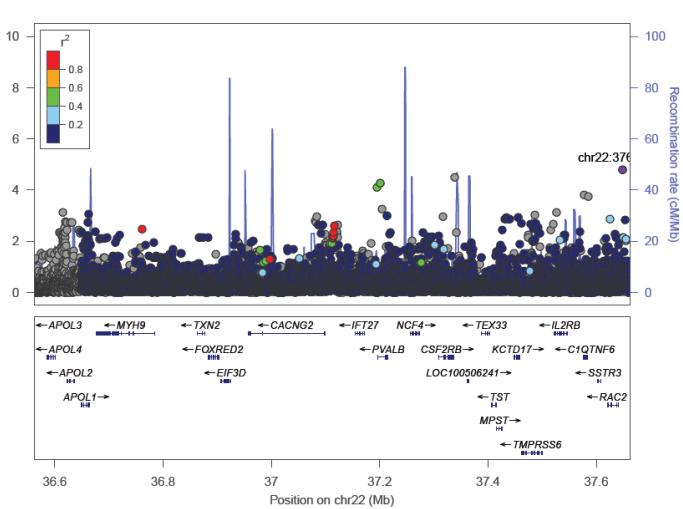
j) Baganda



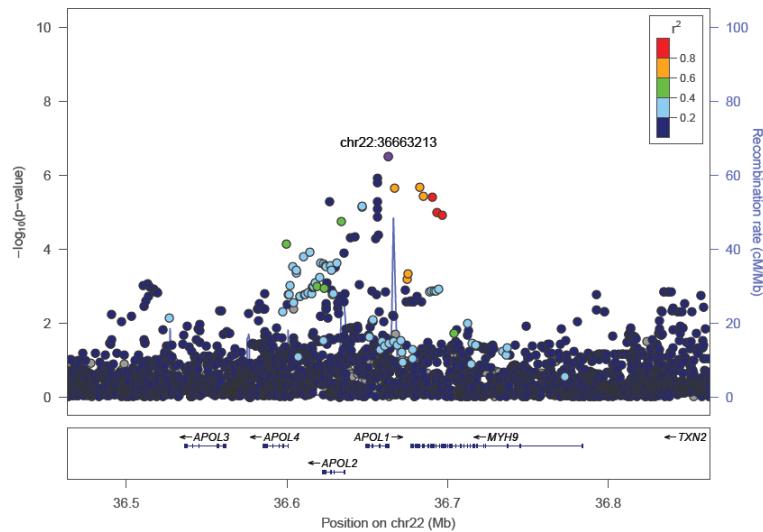
k) Barundi



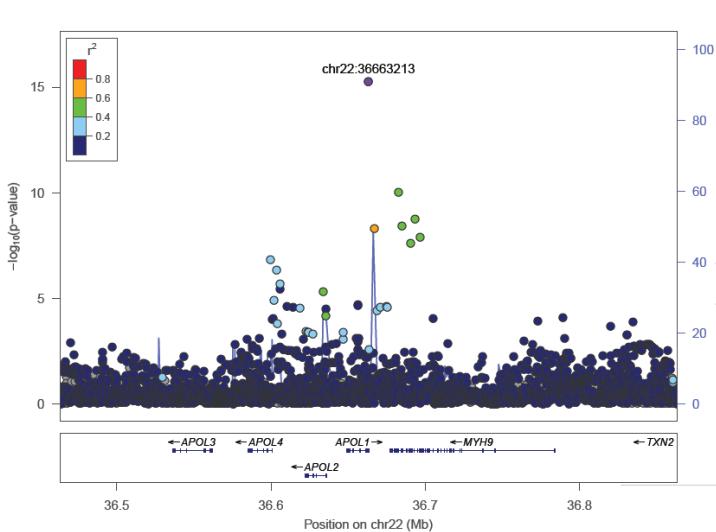
l) Ethiopia



m) Zulu

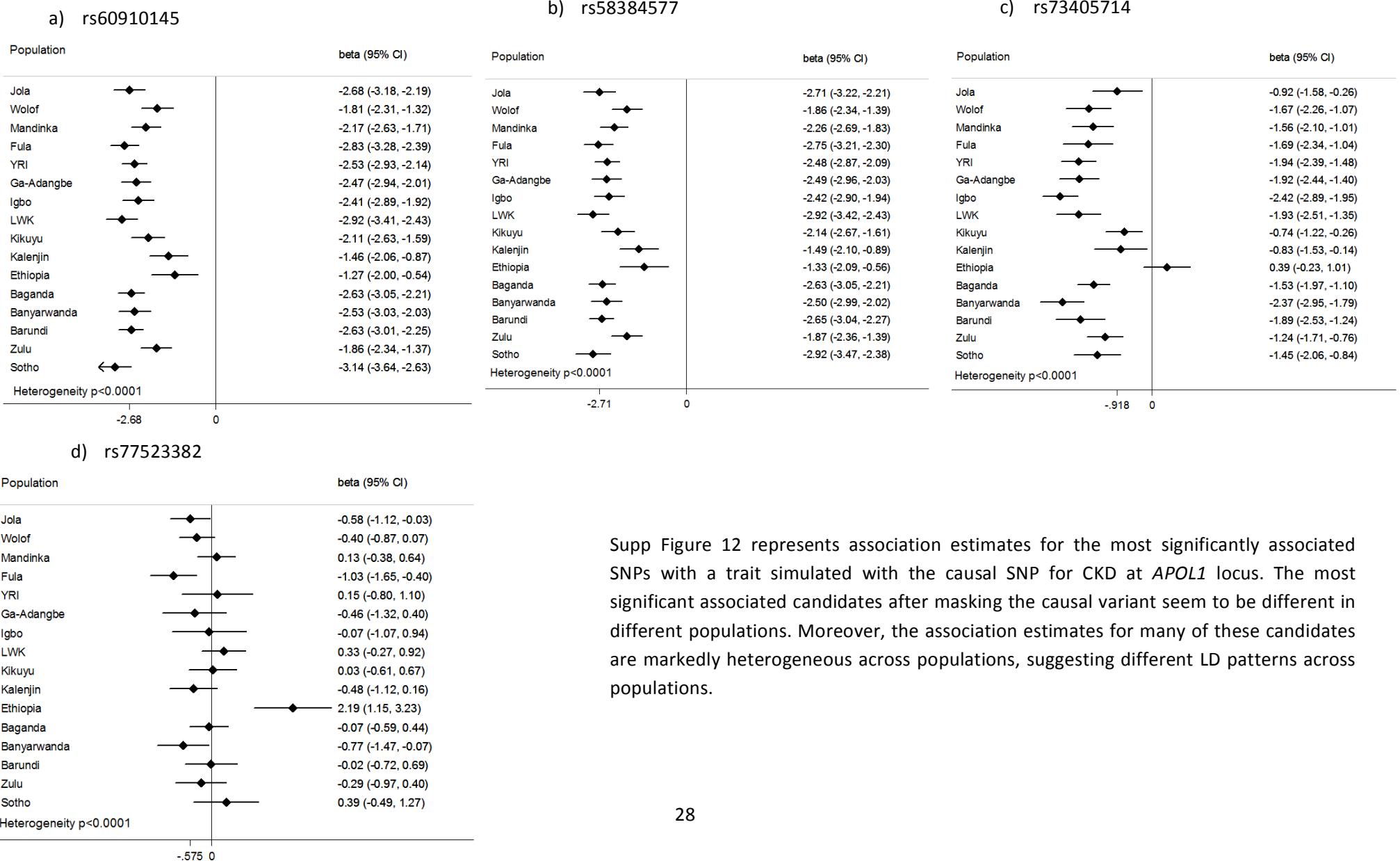


n) Sotho



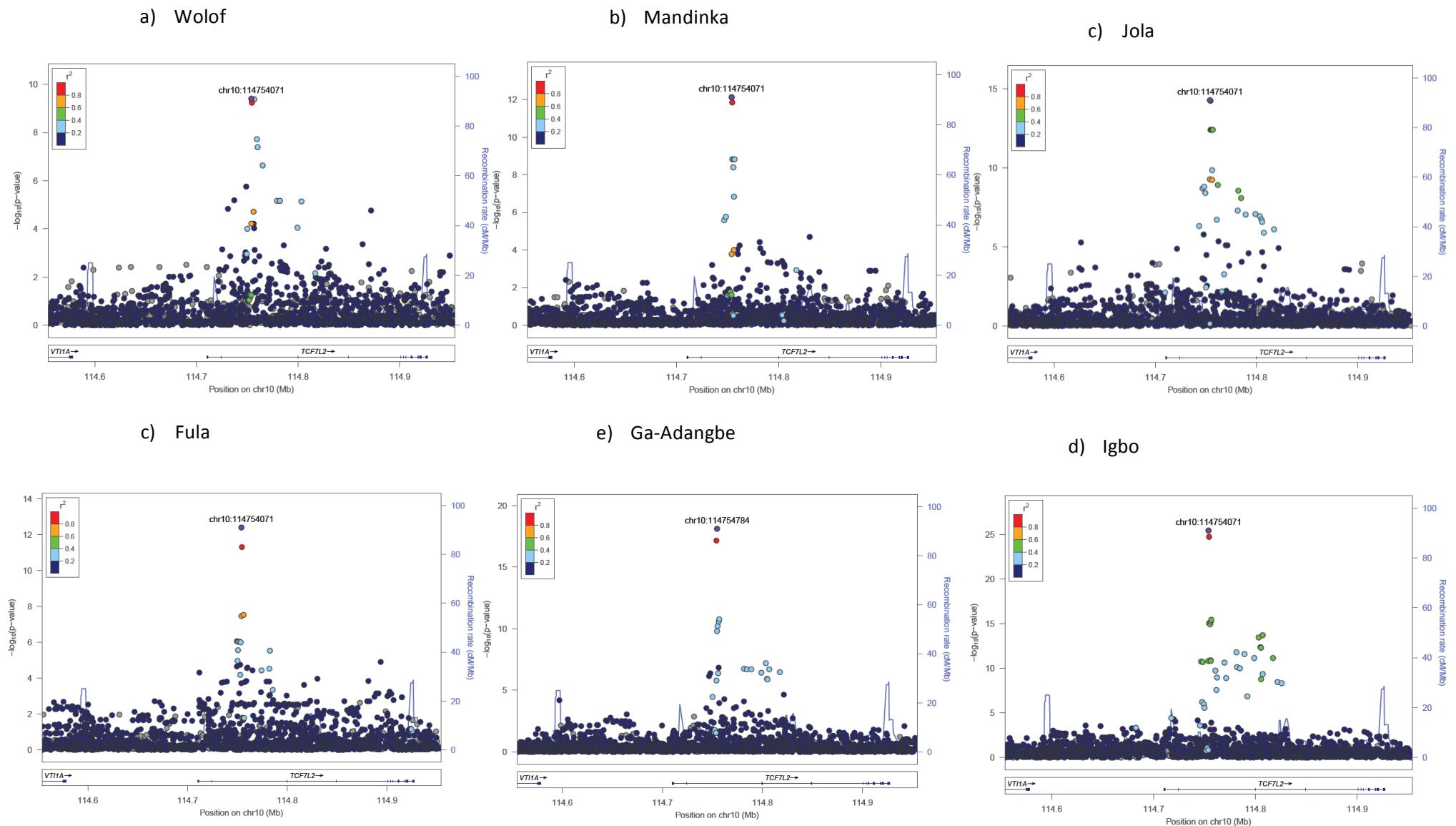
Supp Figure 11 shows the pattern of association for a simulated trait at the *APOL1* locus in different African populations. Regional association plots show very different patterns of association in different populations, suggesting heterogeneous LD structure across African populations in this region.

**Supp. Figure 12: Heterogeneity in effect size at the *APOL1* locus**

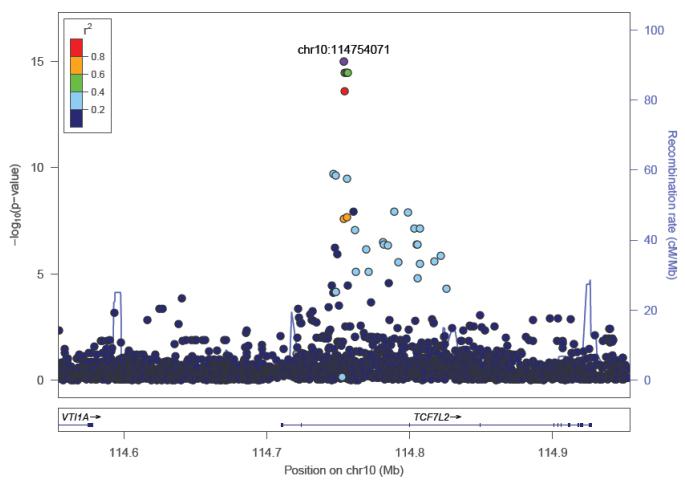


Supp Figure 12 represents association estimates for the most significantly associated SNPs with a trait simulated with the causal SNP for CKD at *APOL1* locus. The most significant associated candidates after masking the causal variant seem to be different in different populations. Moreover, the association estimates for many of these candidates are markedly heterogeneous across populations, suggesting different LD patterns across populations.

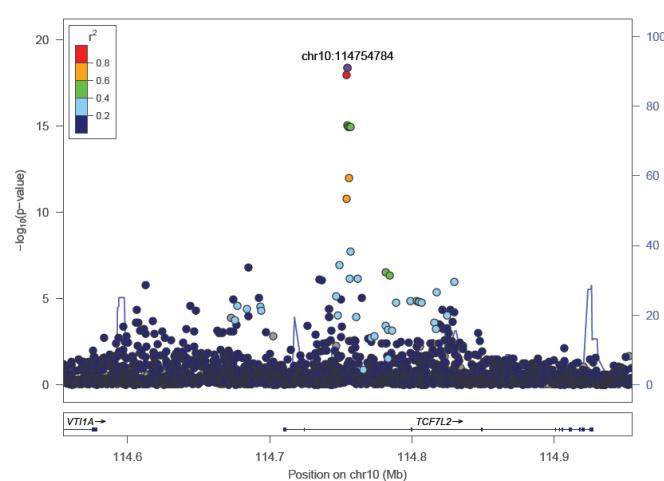
**Supp. Figure 13: Regional association plots and LD structure at the *TCF7L2* locus**



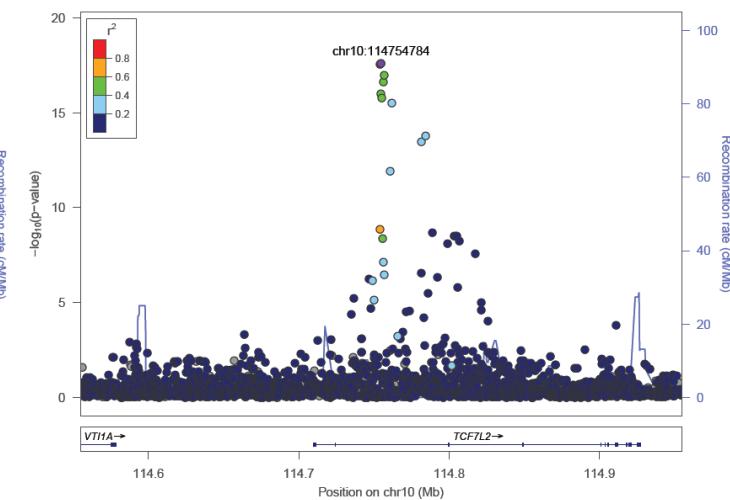
h) YRI



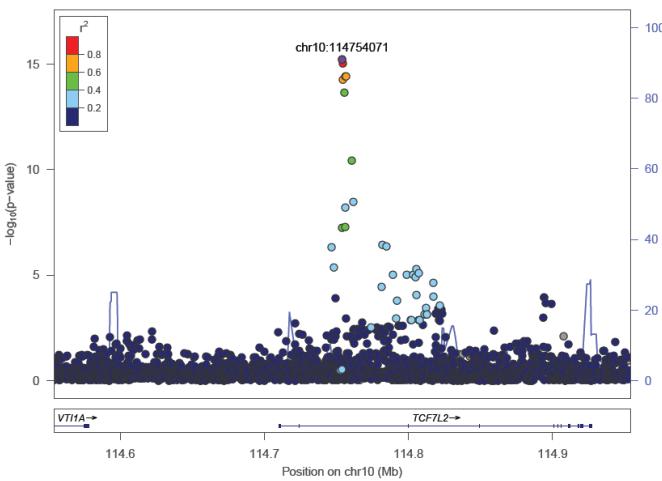
g) Kikuyu



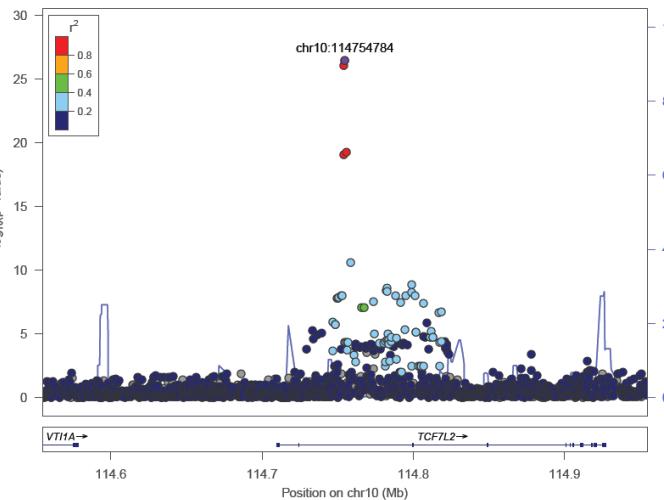
f) Kalenjin



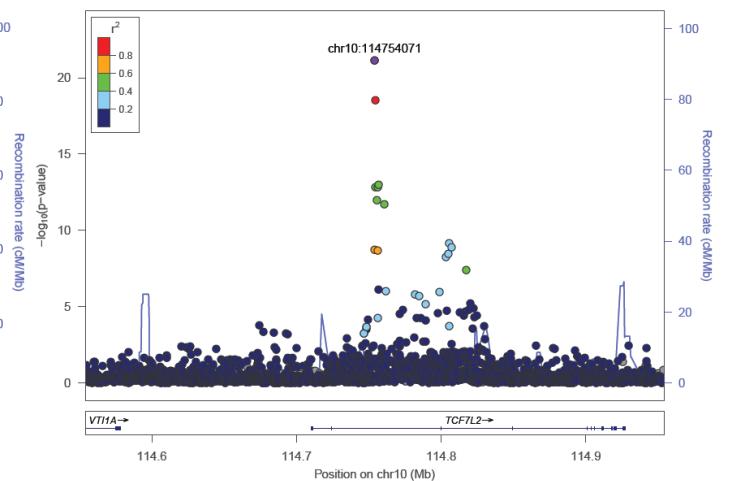
k) LWK



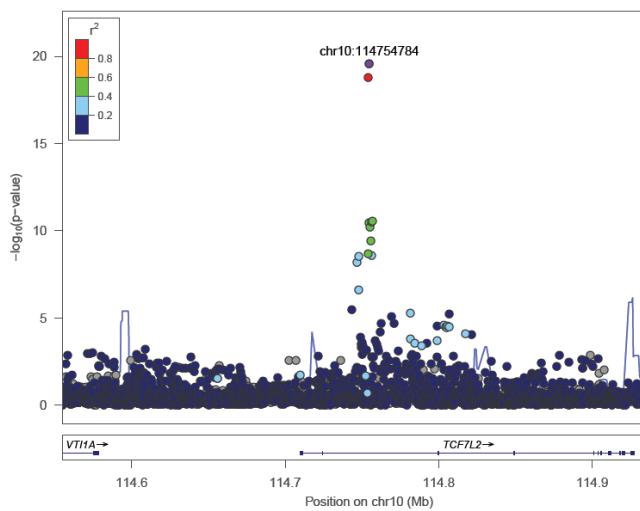
j) Ethiopia



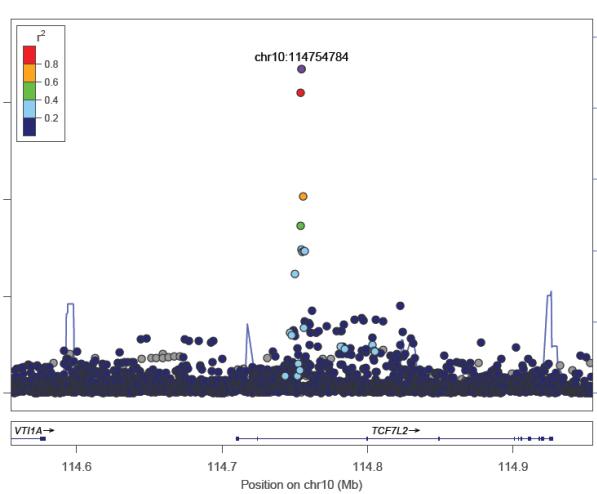
i) Baganda



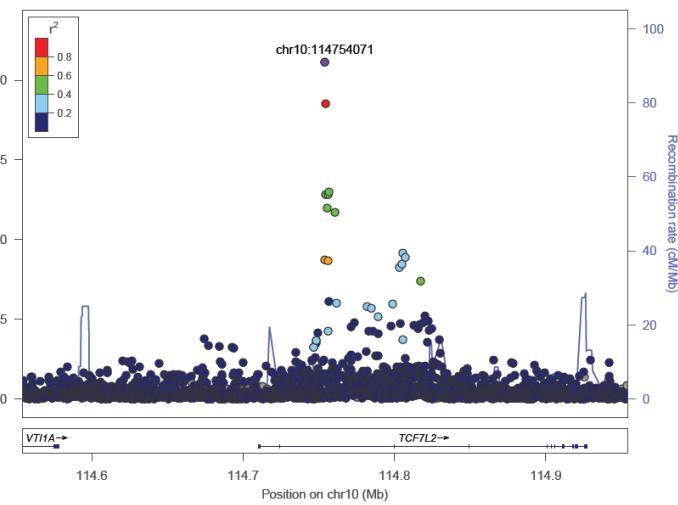
n) Banyawanda



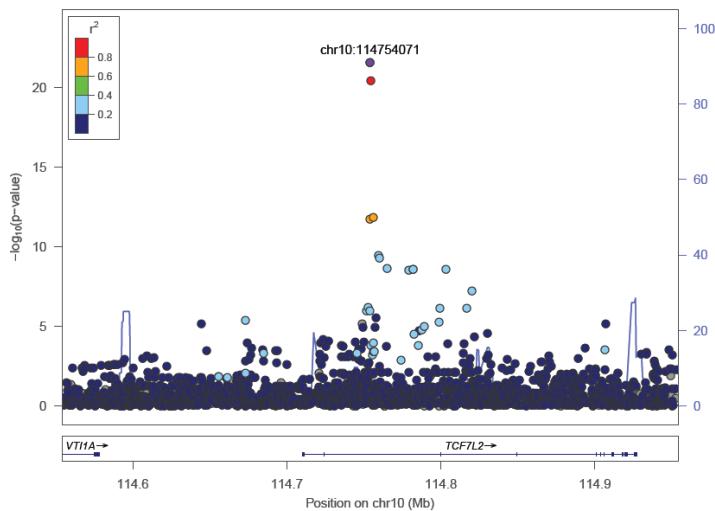
m) Barundi



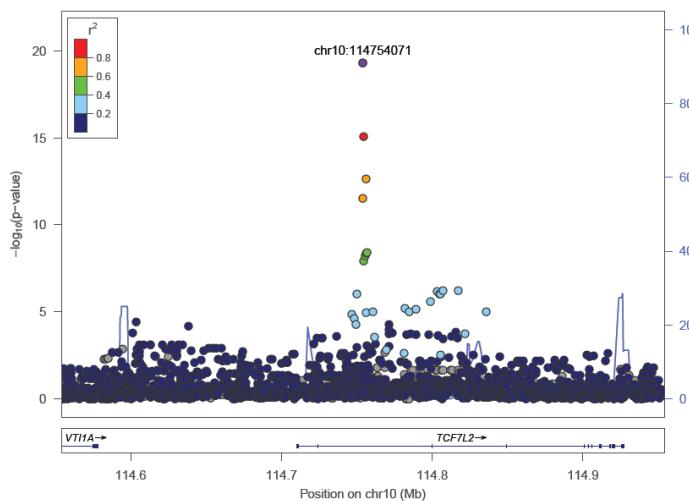
l) Baganda



o) Zulu



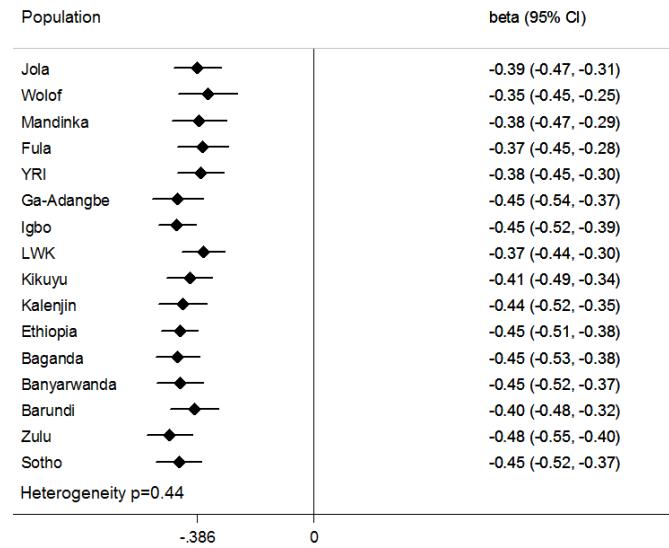
p) Sotho



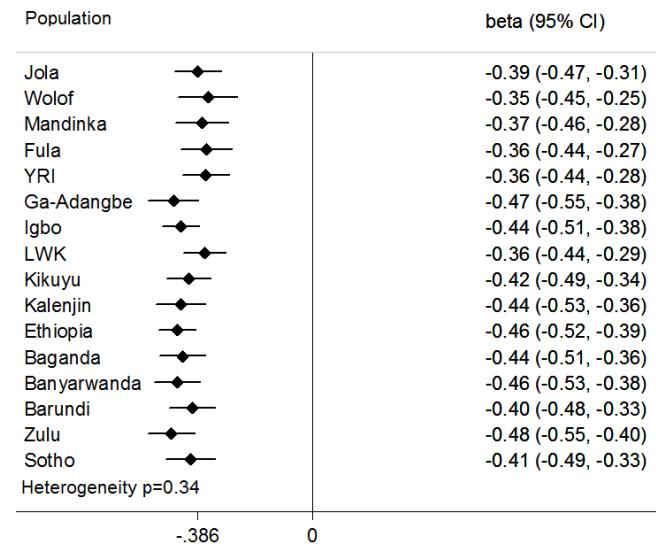
Supp Figure 13 shows the pattern of association for a simulated trait at the *TCF7L2* locus in different African populations. Regional association plots show very similar patterns of across all populations, suggesting homogeneous LD structure across African populations in this region.

**Supp. Figure 14: Homogeneity in test statistics in different African populations at the *TCF7L2* locus**

a) rs34872471

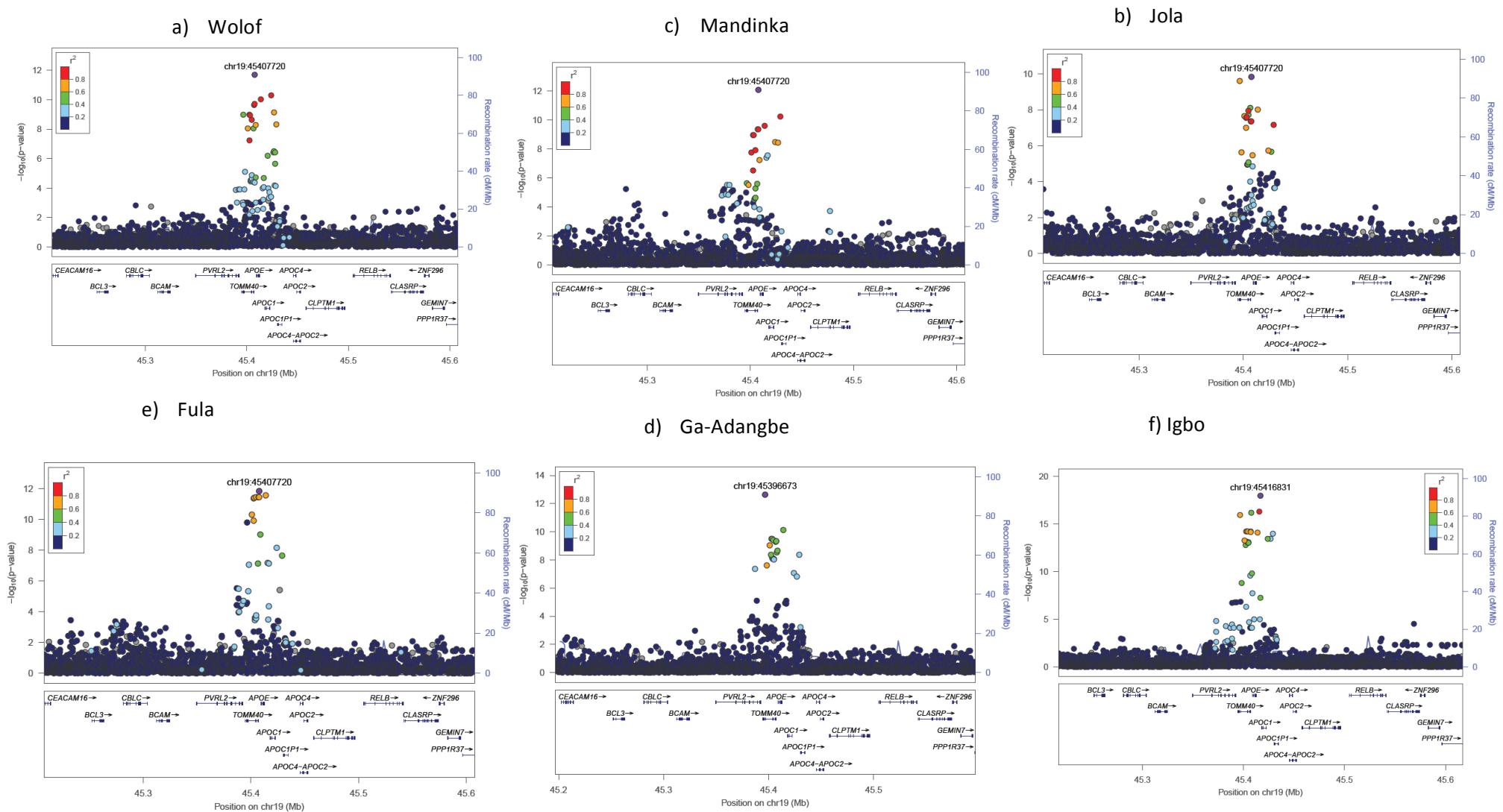


b) rs35198068

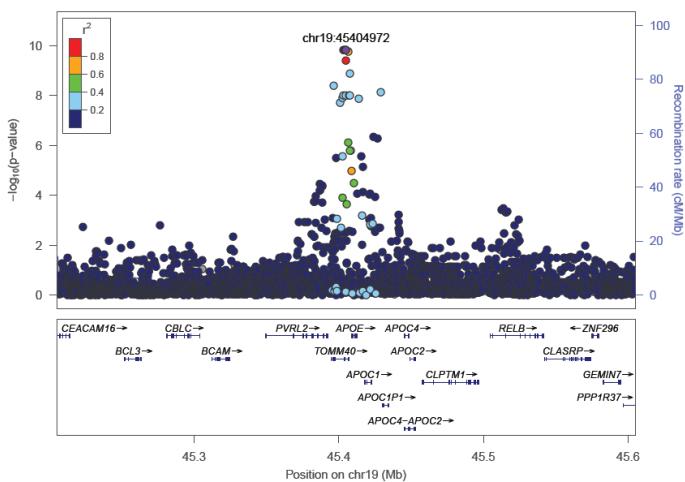


Supp Figure 14 represents association estimates for the most significantly associated SNPs with a trait simulated with the causal *TCF7L2* variant in different populations. Although different top candidates seem to be observed in some populations, the association estimates are homogeneous across all populations, suggesting similar LD patterns across populations, and reproducibility of signals at this locus in association analyses.

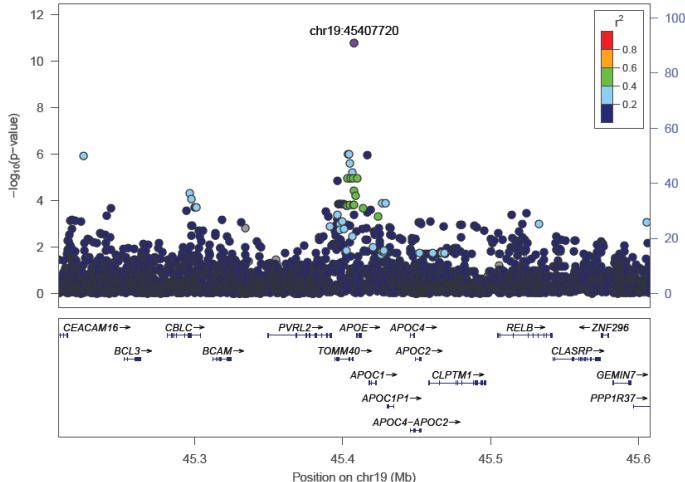
**Supp. Figure 15: Regional association plots and LD structure at the *APOE* locus**



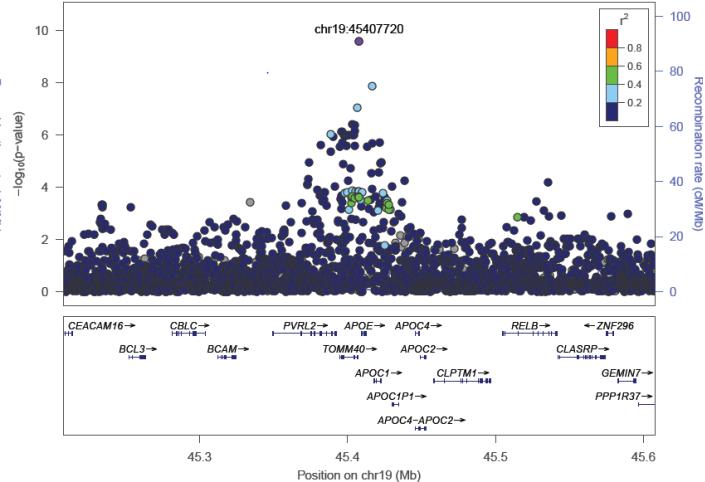
g) YRI



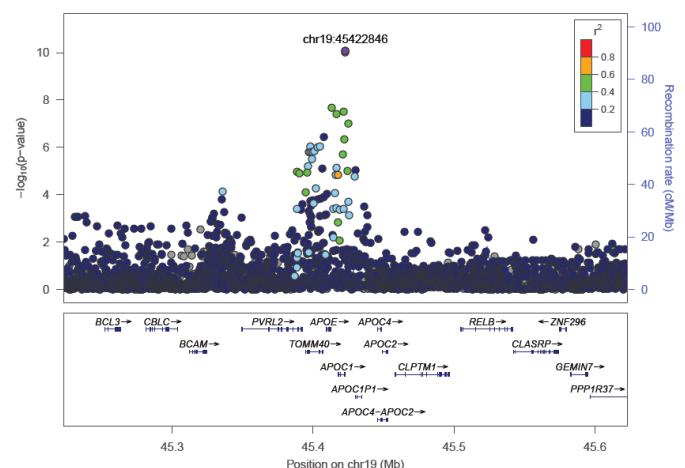
h) LWK



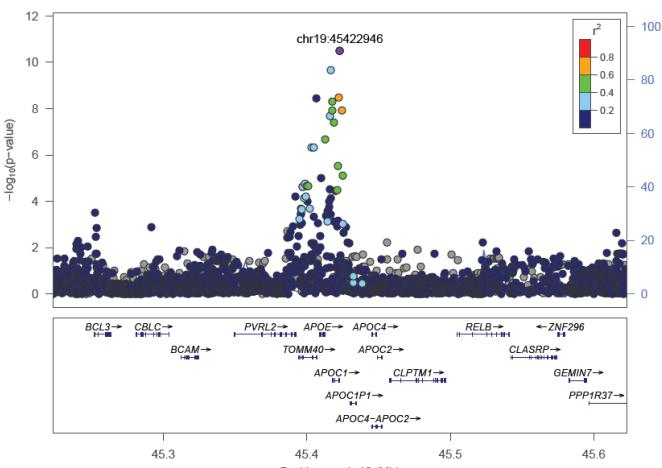
i) Kikuyu



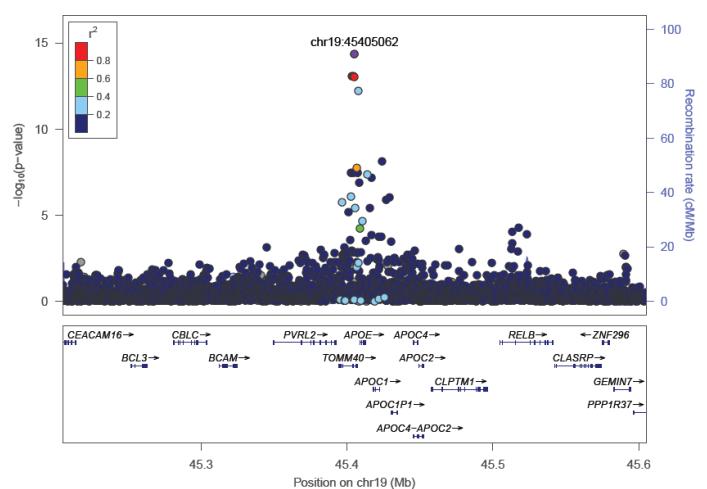
j) Kalenjin



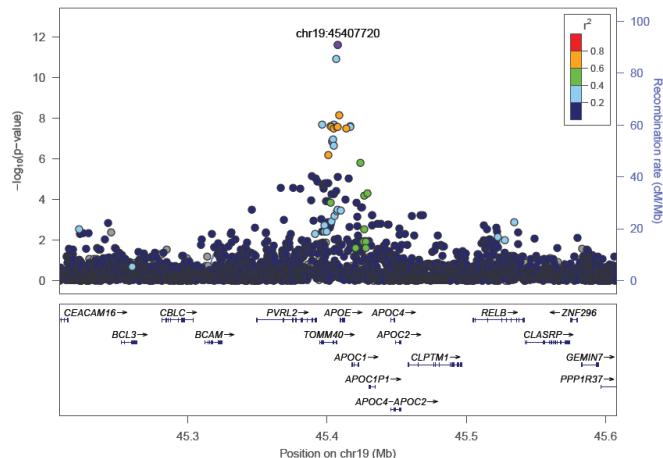
k) Ethiopia



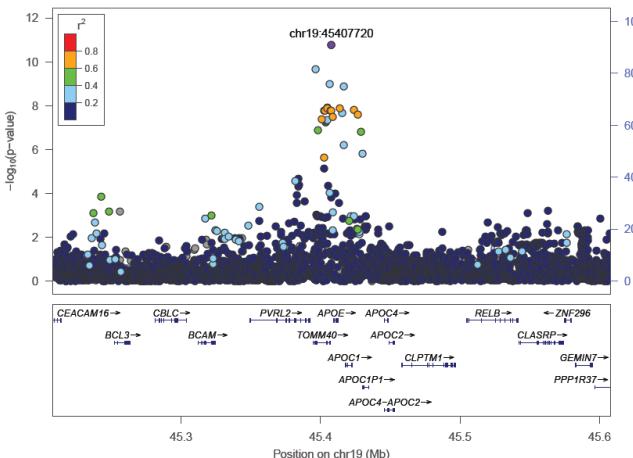
l) Baganda



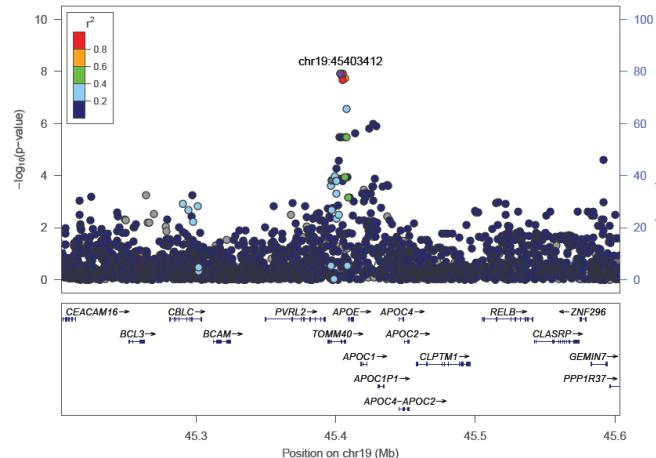
m) Banyarwanda



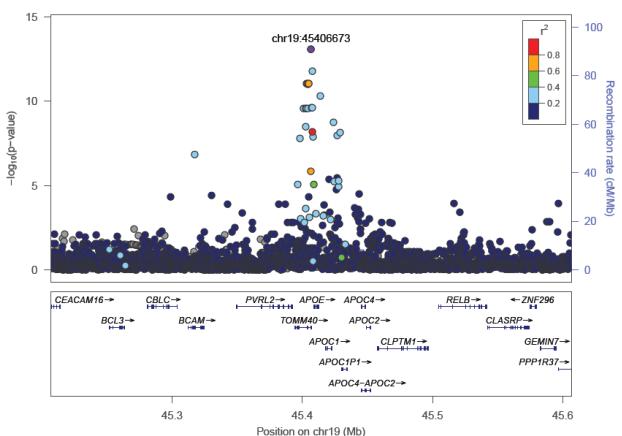
n) Barundi



o) Zulu

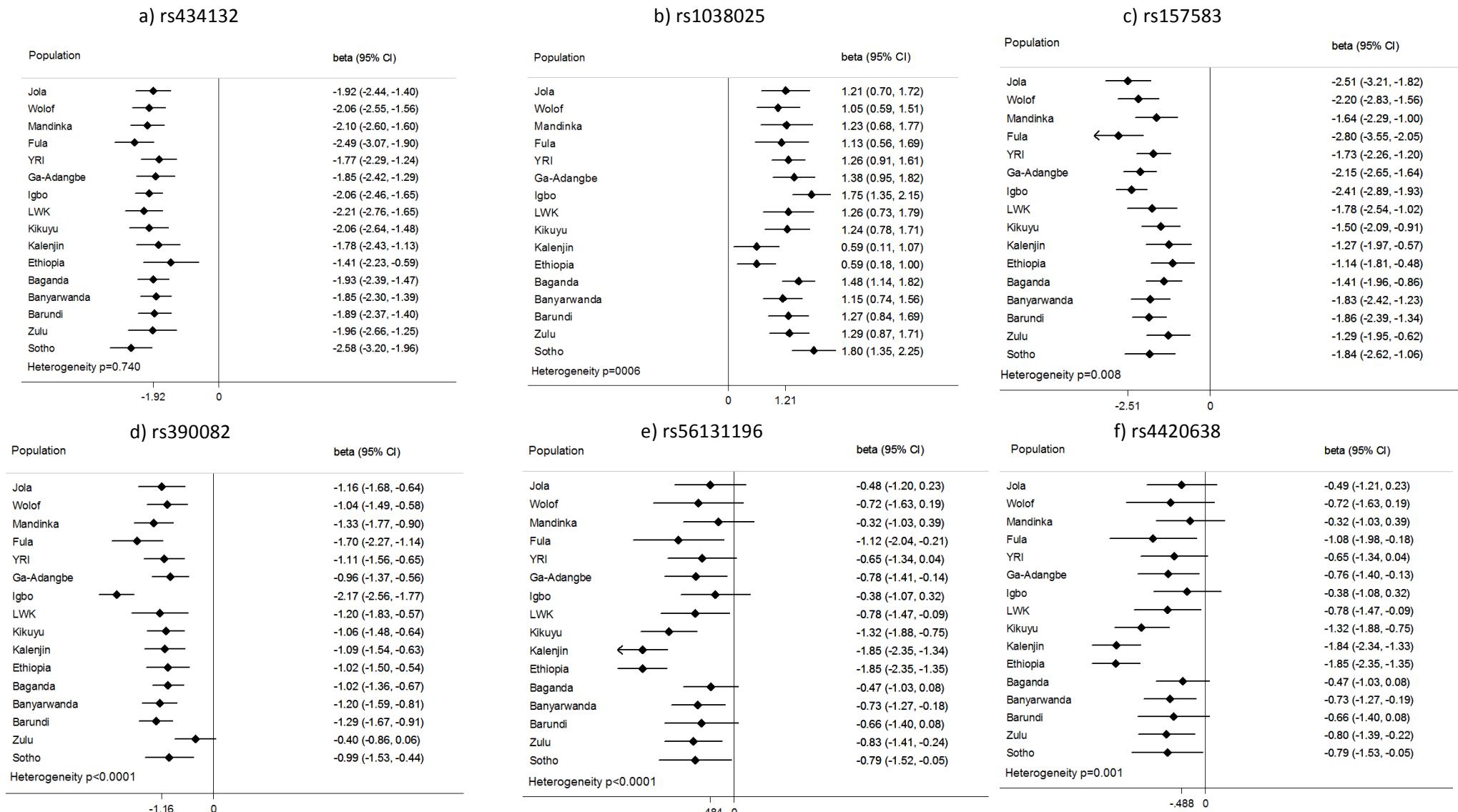


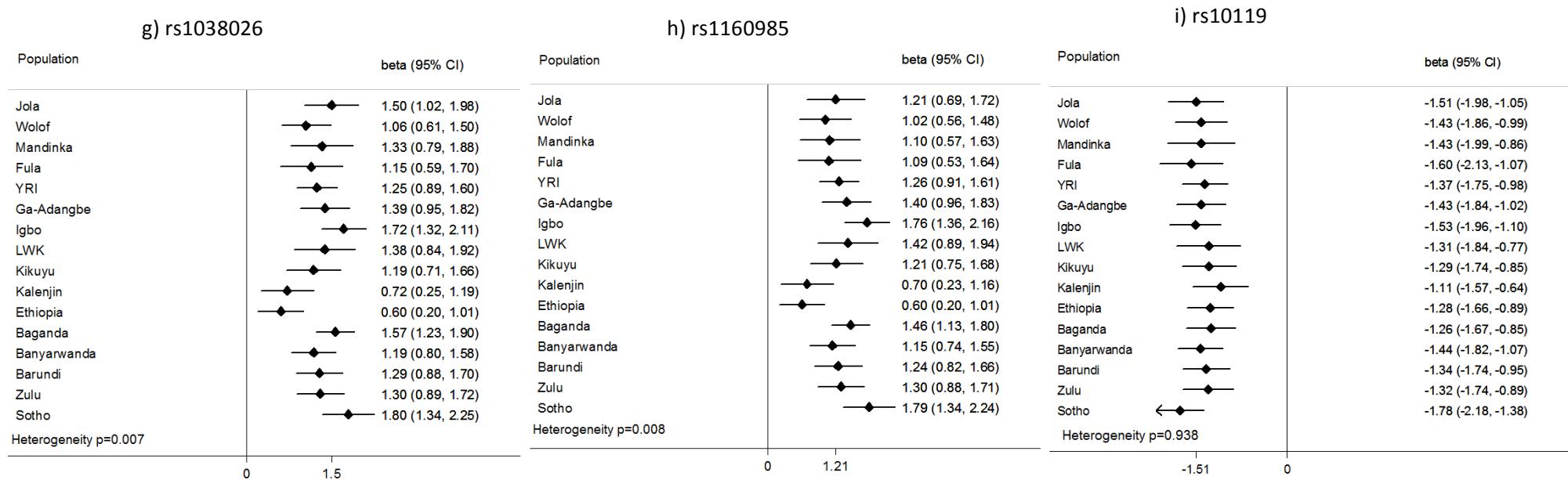
p) Sotho



Supp Figure 15 shows the pattern of association for a simulated trait at the *APOE* locus in different African populations. Regional association plots show different top candidates in different populations with different patterns of association in many populations, suggesting heterogeneous LD structure across African populations in this region.

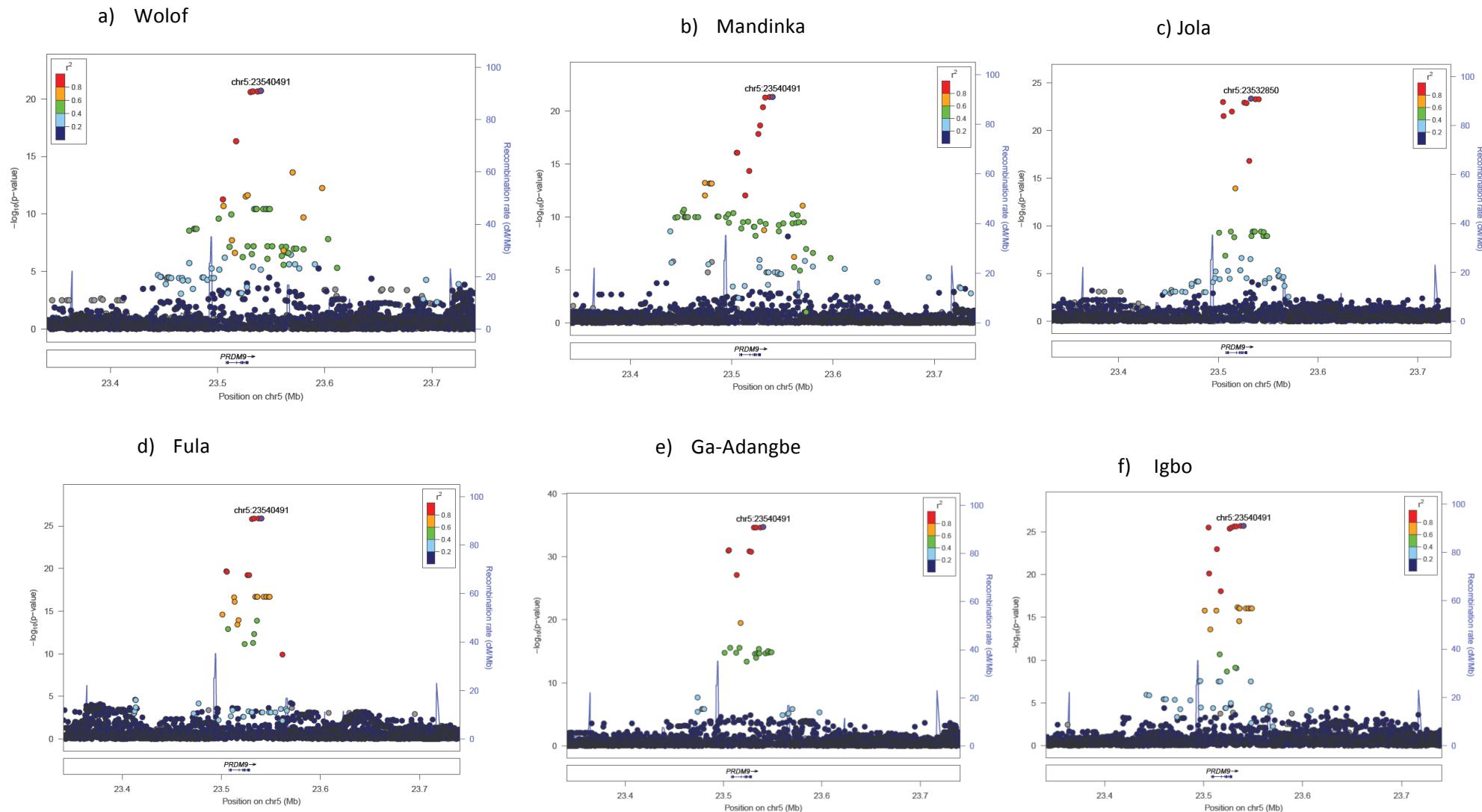
**Supp. Figure 16: Heterogeneity in effect estimates for top candidate associations at the *APOE* locus across African populations**



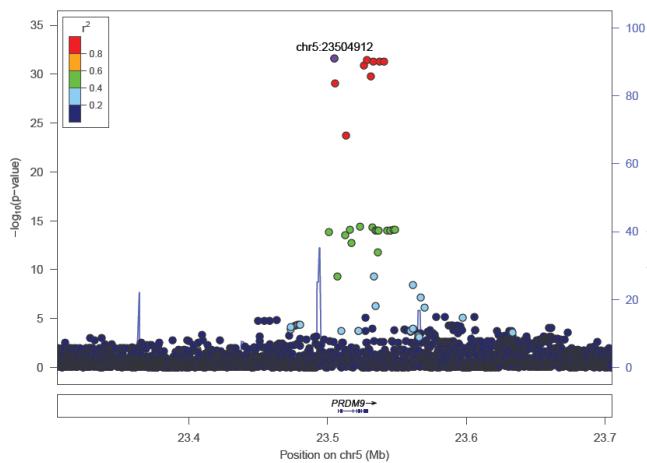


Supp. Figure 16 represents association estimates for the most significantly associated SNPs with a trait simulated with the causal SNP for Alzheimers at the *APOE* locus. The most significant associated candidates after masking the causal variant seem to be very different in different populations. Moreover, the association estimates for many of these candidates are markedly heterogeneous across populations, suggesting different LD patterns across populations.

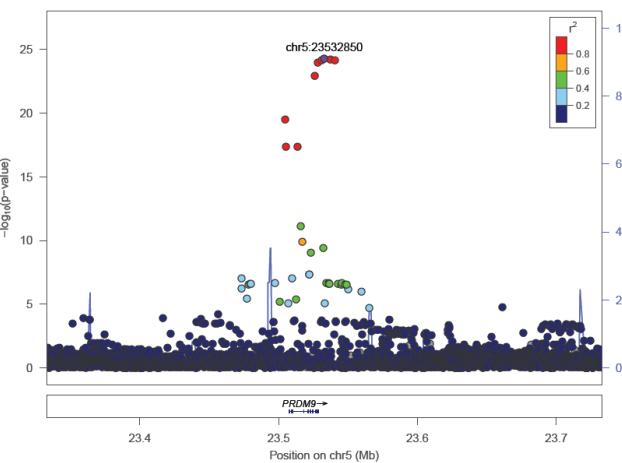
**Supp. Figure 17: Regional association plots and LD structure at the *PRDM9* locus**



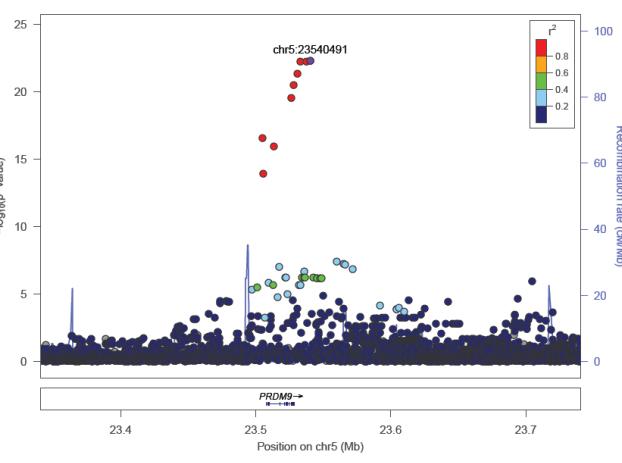
g) YRI



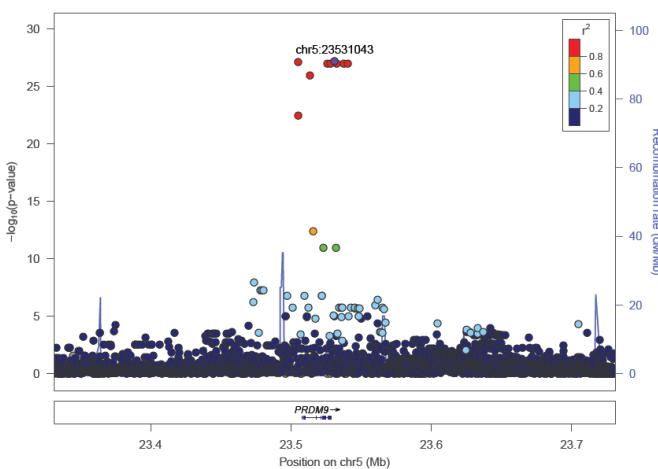
h) Kikuyu



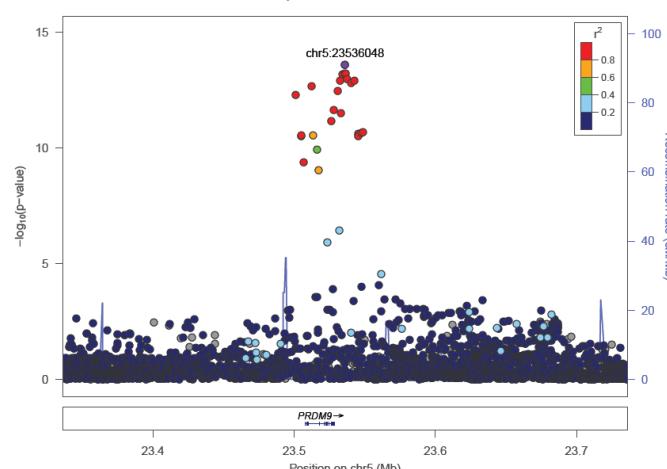
i) Kalenjin



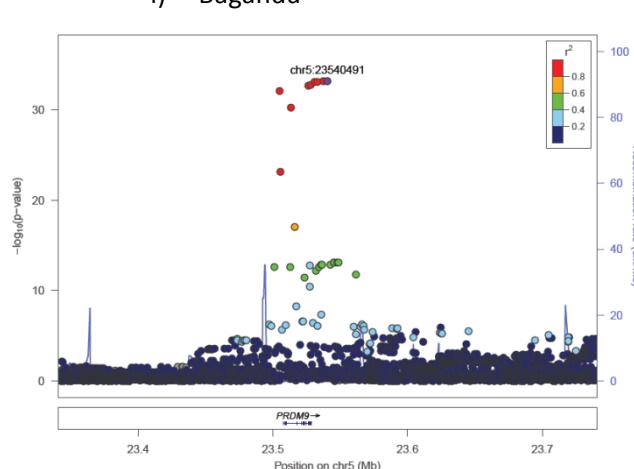
j) LWK

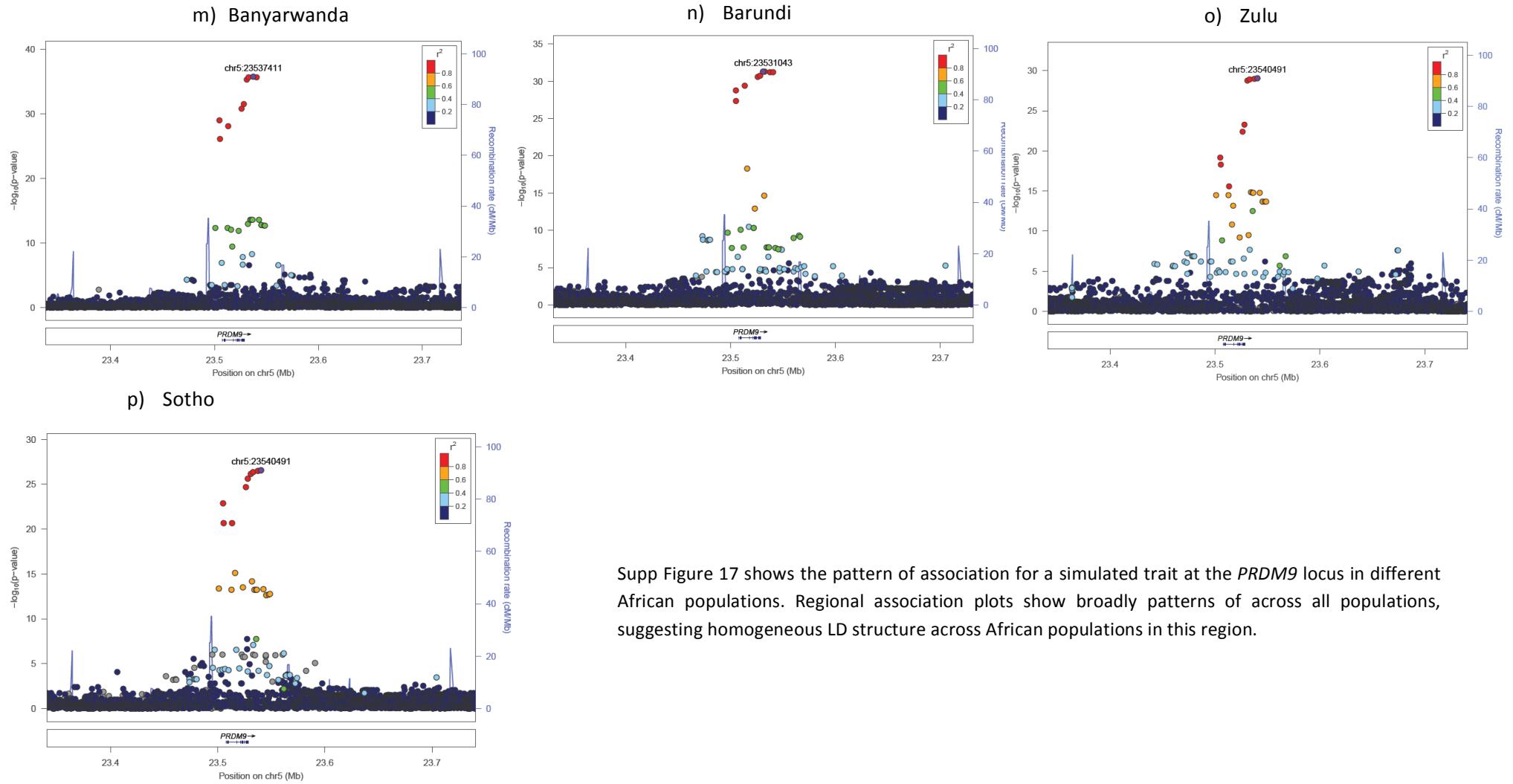


k) Ethiopia



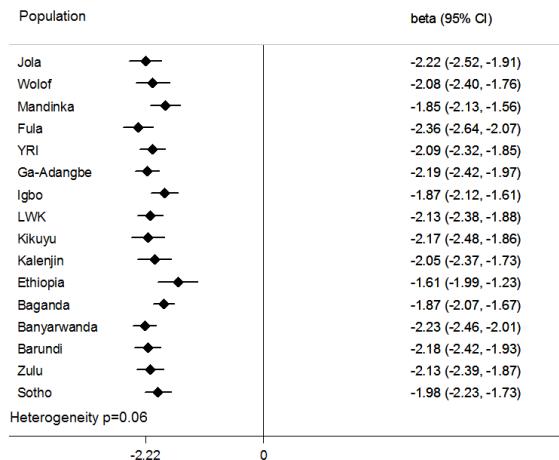
l) Baganda



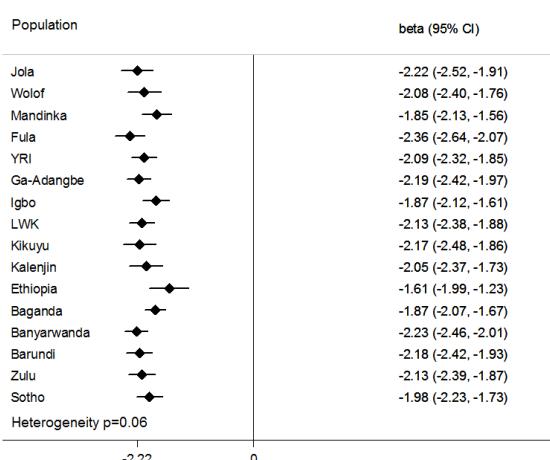


**Supp. Figure 18: Heterogeneity in effect estimates for *PRDM9* across African populations**

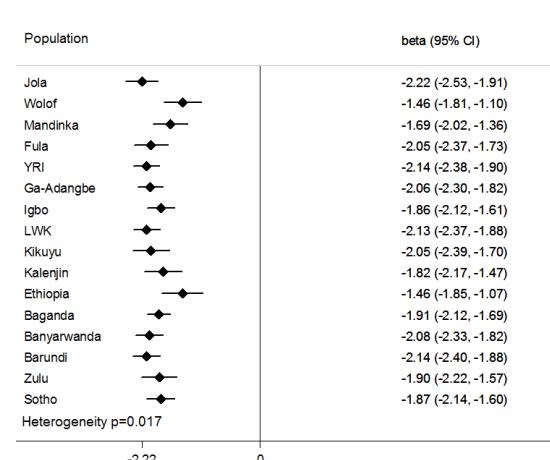
a) rs57423488



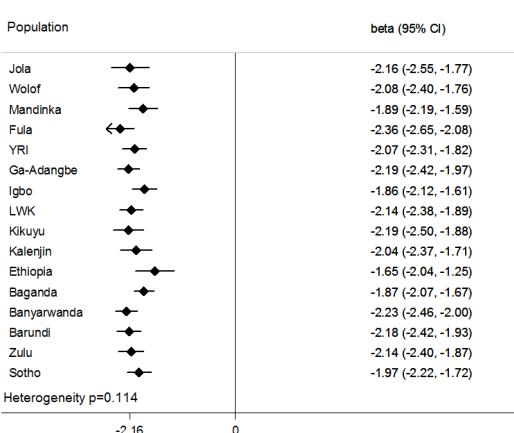
b) rs2914277



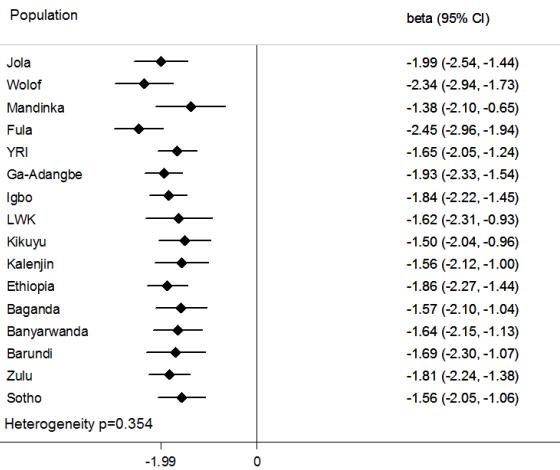
c) rs2973613



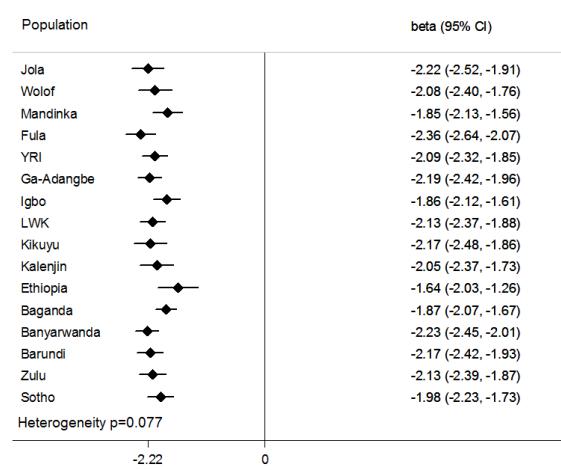
d) rs2914280



e) rs55873106



f) chr5:23537411:l



Supp Figure 18 represents association estimates for the most significantly associated SNPs with a trait simulated at the PRDM9 locus in different populations. Although different top candidates seem to be observed in many populations, the association estimates for most are broadly homogeneous across populations, suggesting similar LD patterns across populations, and reproducibility of signals at this locus in association analyses.